BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

As rescanning documents *will not* correct images, please do not report the images to the Image Problem Mailbox.

Pharmaceutical Biotechnology

Series Editor: Ronald T. Borchardt
The University of Kansas
Lawrence, Kansas

Recent volumes in this series:

Volume 2 STABILITY OF PROTEIN PHARMACEUTICALS, Part A: Chemical and Physical Pathways of Protein Degradation
Edited by Tim J. Ahern and Mark C. Manning

Volume 3 STABILITY OF PROTEIN PHARMACEUTICALS, Part B: *In Vivo* Pathways of Degradation and Strategies for Protein Stabilization
Edited by Tim J. Ahern and Mark C. Manning

Volume 4 BIOLOGICAL BARRIERS TO PROTEIN DELIVERY
Edited by Kenneth L. Audus and Thomas J. Raub

Volume 5 STABILITY AND CHARACTERIZATION OF PROTEIN AND PEPTIDE DRUGS: Case Histories
Edited by Y. John Wang and Rodney Pearlman

Volume 6 VACCINE DESIGN: The Subunit and Adjuvant Approach
Edited by Michael F. Powell and Mark J. Newman

Volume 7 PHYSICAL METHODS TO CHARACTERIZE PHARMACEUTICAL PROTEINS
Edited by James N. Herron, Wim Jiskoot, and Daan J. A. Crommelin

Volume 8 MODELS FOR ASSESSING DRUG ABSORPTION AND METABOLISM
Edited by Ronald T. Borchardt, Philip L. Smith, and Glynn Wilson

Volume 9 FORMULATION, CHARACTERIZATION, AND STABILITY OF PROTEIN DRUGS: Case Histories
Edited by Rodney Pearlman and Y. John Wang

Formulation, Characterization, and Stability of Protein Drugs

Case Histories

Edited by

Rodney Pearlman
Megabios Corporation
Burlingame, California

and

Y. John Wang
Scios Nova, Inc.
Mountain View, California

Plenum Press • New York and London

Library of Congress Cataloging-in-Publication Data

On file

ISBN 0-306-45332-0

A Division of Plenum Publishing Corporation
233 Spring Street, New York, N. Y. 10013

10 9 8 7 6 5 4 3 2 1

Printed in the United States of America

To Jessica and Lynn

Steinetz, B. G., Beach, V. L., and Kroc, R. L., 1969, Bioassay of relaxin, in: *Methods in Hormone Research*, 2nd ed. (R. I. Dorfman, ed.), Academic Press, New York, pp. 481–513.

Steinetz, B. G., O'Byrne, E. M., Butler, M. C., and Hickman, L. B., 1982, Hormonal regulation of the connective tissue of the symphysis pubis, in: *Biology of Relaxin and Its Role in the Human* (M. Bigazzi, F. C. Greenwood, and F. Gaspari, eds.), Excerpta Medica, Princeton, NJ, p. 71.

Stults, J. T., Bourell, J. H., Canova-Davis, E., Ling, V. T., Laramee, G. R., Winslow, J. W., Griffin, P. R., Rinderknecht, E., and Vandlen, R. L., 1990, Structural characterization by mass spectrometry of native and recombinant human relaxin, *Biomed. Environ. Mass Spectrom.* **19**:655–664.

Unemori, E. L., and Amento, E. P., 1990, Relaxin modulates synthesis and secretion of procollagenase and collagen by human dermal fibroblasts, *J. Biol. Chem.* **265**:337–342.

Unemori, E. L., Bauer, E. A., and Amento, E. P., 1992, Relaxin alone and in conjunction with interferon-gamma decreases collagen synthesis by cultured human scleroderma fibroblasts, *J. Invest. Dermatol.* **99**:337–342.

Unemori, E. N., Beck, S. L., Lee, W. P., Xu, Y., Siegel, M., Keller, G., Liggit, H. D., Bauer, E. A., and Amento, E. P., 1993, Human relaxin decreases collagen accumulation *in vivo* in two rodent models of fibrosis, *J. Invest. Dermatol.* **101**:280–285.

Ward, D. G., Thomas, G. R., and Cronin, M. J., 1992, Relaxin increases rat heart rate by a direct action on the cardiac atrium, *Biochem. Biophys. Res. Commun.* **186**:999–1005.

Weiss, G., 1989, Relaxin in the male, *Biol. Reprod.* **40**:197–200.

Winslow, J. W., Shih, A., Bourell, J. H., Weiss, G., Reed, B., Stults, J. T., and Goldsmith, L. T., 1992, Human seminal relaxin is the same product gene as human luteal relaxin, *Endocrinology* **130**:2660–2668.

Wood, S. P., Blundell, T. L., Lazarus, N. R, and Neville, W. J., 1975, The relation of conformation and association of insulin to receptor binding; x-ray and circular-dichroism studies on bovine and hystricomorph insulins, *Eur. J. Biochem.* **55**:531–542.

6

Interferon-β-1b (Betaseron®): A Model for Hydrophobic Therapeutic Proteins

Leo S. Lin, Michael G. Kunitani, and Maninder S. Hora

1. INTRODUCTION

Interferon-β-1b is a form of interferon-β (IFN-β) which has shown biological activity in a variety of *in vitro* and *in vivo* systems. IFN-β belongs to a class of proteins known as interferons (IFNs). Interferons were originally classified based on the cell type from which they were derived. Thus, the three major classes of IFNs were designated as leukocyte-, fibroblast-, and immune-interferon as these species were predominantly synthesized by leukocytes, fibroblasts, and T-lymphocytes, respectively (Pestka, 1983; Zoon, 1987). With our increasing knowledge of IFN structure and function, the nomenclature of IFN has also evolved. Today, the three major classes of IFN are referred to as IFN-α, IFN-β, and IFN-γ. Human IFN-α and -β, are approximately 30% similar at their primary amino acid sequence level, while IFN-γ is similar to neither. It is also believed that IFN-α and IFN-β bind to the same IFN receptor while there is a separate receptor for IFN-γ (Faltynek and Baglioni, 1984).

Natural human IFN-β is a glycoprotein with an approximate molecular weight of 23,000 Daltons. Correctly engineered recombinant, nonglycosylated, IFN-β species (molecular weight 18,500 Daltons) display the same biological effects as the native molecule. The IFN-β protein has been associated with a variety of antiviral

Leo S. Lin and Michael G. Kunitani • Department of Analytical Development, Chiron Corporation, Emeryville, California 94608. *Maninder S. Hora* • Department of Formulation Development, Chiron Corporation, Emeryville, California 94608.

Formulation, Characterization, and Stability of Protein Drugs, Rodney Pearlman and Y. John Wang, eds., Plenum Press, New York, 1996.

(Kerr and Stark, 1992; Soike, 1987), antiproliferative (Arabje *et al.*, 1993), antiinfective (Kirchner, 1986), and immunomodulating (Reiter, 1993; Murray, 1992) activities. A brief history of the interferons, including IFN-β, has been discussed by Dianzani and Dolei (1984).

2. MOLECULAR BIOLOGY AND PROTEIN CHEMISTRY

The human IFN-β gene was cloned and expressed in a variety of host systems under the control of different promoter systems. A production strain of the bacterium *Escherichia coli* (*E. coli*) harboring a recombinant plasmid containing the human IFN-β gene and capable of expressing a part of its cellular proteins as recombinant IFN-β was introduced into well-controlled fermentation processes. Recombinant human IFN-β (rhIFN-β) was extracted from cells and purified by a series of column chromatographic and other steps (Mark *et al.*, 1984). The resulting product, purified to >95% purity as determined by sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE), displayed a specific activity that was about 10-fold less than that of IFN-β produced from cultured human fibroblast cells. It was also found that most of the IFN-β protein existed in its covalent-linked dimeric and higher oligomeric forms in *E. coli*. Furthermore, the purified rhIFN-β exhibited loss of purity and potency over time (Mark *et al.*, 1984).

IFN-β has three cysteine residues, located at amino acid positions 17, 31, and 141. One or more of these cysteines could be involved in intermolecular disulfide bridging, resulting in the formation of inactive dimers and oligomers. Likewise, the three cysteines may also interact randomly within each molecule, resulting in three types of molecular species in the cell, each one with one of the three possible intramolecular disulfide bridges. It was postulated that only one of these forms may resemble the native conformation and retain biological activity. Both these possibilities could together result in the formation of inactive monomers and oligomers in the cell. If the sulfhydryls were responsible for the lower specific activity of the IFN-β protein, then removal of one of the cysteines would allow only one unique intramolecular disulfide bridge formation, leaving no free-sulfhydryl group to generate dimers or oligomers. Therefore, it was sought to eliminate one of the three cysteine residues by site-specific mutagenesis of the IFN-β gene, whereby one of the codons for cysteine is changed to that of serine. Serine was chosen as a replacement for cysteine because the two amino acids differ by only a single atom: the cysteine residue has a sulfur atom that is replaced by an oxygen atom in the serine residue. Cys-141 of the IFN-β molecule was known to be required for biological activity (Shepard *et al.*, 1981). By analogy with the IFN-α molecules in which a -S-S- bond is formed between Cys-29 and Cys-138 (Wetzel *et al.*, 1981), it was thought that the Cys-141 of IFN-β could be involved in a disulfide bridge with Cys-31, leaving a free and reactive thiol group on Cys-17. The Cys-17 residue was therefore chosen for

replacement with serine. A schematic diagram showing the primary sequence of IFN-β_{ser17} is presented in Fig. 1.

The biological activities of IFN-β_{cys17} and IFN-β_{ser17} were compared in a virus yield reduction assay. The purified IFN-β_{cys17} had a specific antiviral activity of 3×10^7 units/mg. In contrast, the purified IFN-β_{ser17} exhibited a specific activity of 2×10^8 units/mg, comparable to that of purified native IFN-β (Derynick *et al.*, 1980). The biological activity of purified preparations of IFN-β_{cys17} and IFN-β_{ser17} were compared in a number of studies. Figure 2 illustrates the activity profile of the two IFN species stored at −70°C. The activity of IFN-β_{ser17} remained unchanged over a period of 150 days, while IFN-β_{cys17} lost a significant amount of its antiviral activity in 75 days. In addition, when these preparations were analyzed by nonreducing SDS-PAGE, a significant amount of dimers and oligomers could be detected in the IFN-β_{cys17} sample but not in the IFN-β_{ser17} preparation (Mark *et al.*, 1984). These data demonstrate that substitution of the cysteine residue at position 17 in the IFN-β with a serine residue prevents the formation of incorrect disulfide bonds resulting in a stable and bioactive rhIFN-β molecule. The IFN-β_{ser17} mutein was further developed as Betaseron® by Cetus Corporation, now Chiron Corporation in collaboration with Berlex Biosciences. The IFN-β_{ser17} molecule has been assigned an USAN name of IFN-β-1b.

Similar to the situation with human IFN-β, the Cys-31–141 disulfide bond is also important for biological activity of recombinant murine IFN-β synthesized in *E. coli* (Day *et al.*, 1992).

3. PRECLINICAL AND CLINICAL APPLICATIONS OF IFN-β

3.1. Preclinical Studies

The pharmacokinetics and antiviral activity of IFN-β_{ser17} (Betaseron®) were evaluated in an African green monkey model. This animal model has been successfully used for the evaluation of efficacy and pharmacokinetics of antiviral agents (Soike *et al.*, 1987, 1990). IFN-β_{ser17} was administered by the intravenous, intramuscular, and subcutaneous routes. Following i.v. administration, mean clearance, steady-state volume of distribution and terminal half-life values were 0.36 ± 0.08 liters/hr-kg, 0.65 ± 0.09 liters/kg, and 1.9 ± 0.43 hr, respectively. Bioavailability values for IFN-β_{ser17} delivered by the intramuscular and subcutaneous routes were determined to be 51% and 31%, respectively. Despite only 30–50% bioavailability by these non-i.v. routes, antiviral activity was comparable for i.v., i.m., and s.c. administration of 1×10^6 IU/kg of IFN-β_{ser17} twice daily (Chiang *et al.*, 1993). These studies also indicated that higher doses of the protein resulted in increases of the area under the serum concentration-time curve and of its antiviral efficacy. Finally, these studies demonstrated that significant accumulation of IFN-β_{ser17} in serum occurred with

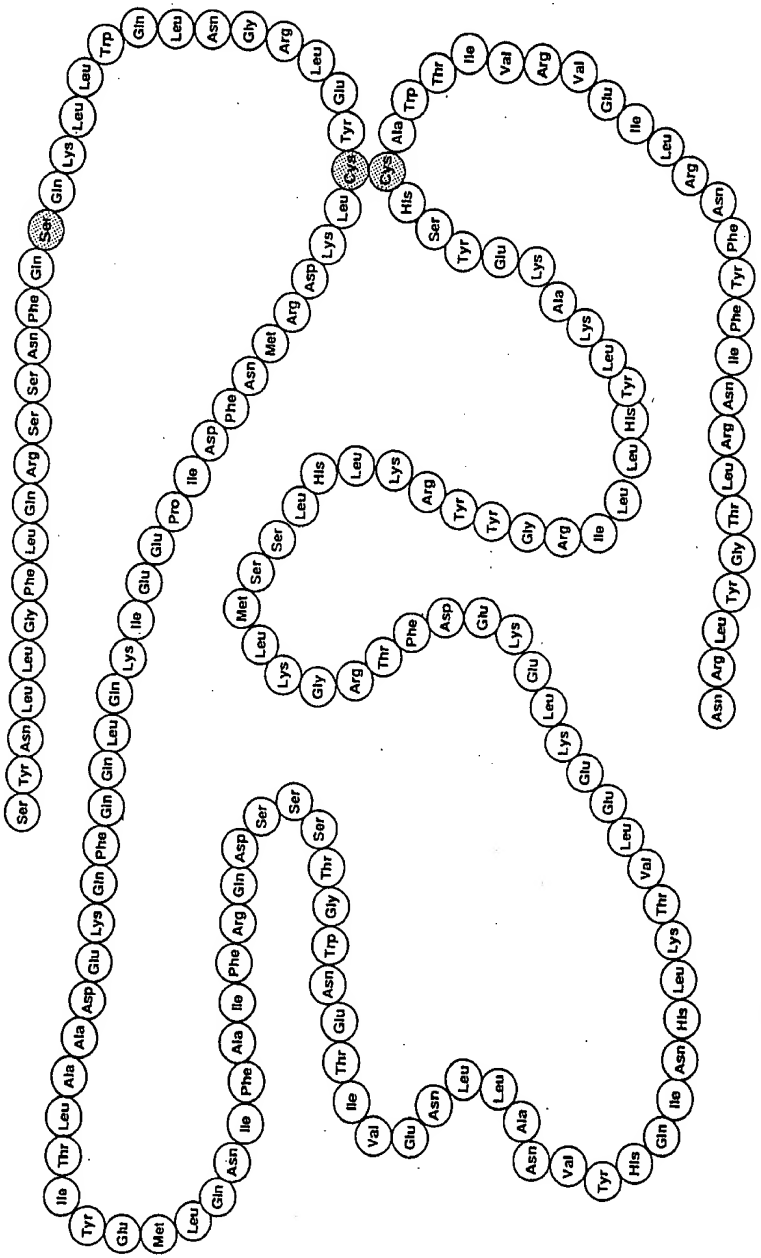


Figure 1. The primary amino acid sequence of recombinant human interferon-β_{ser17}.

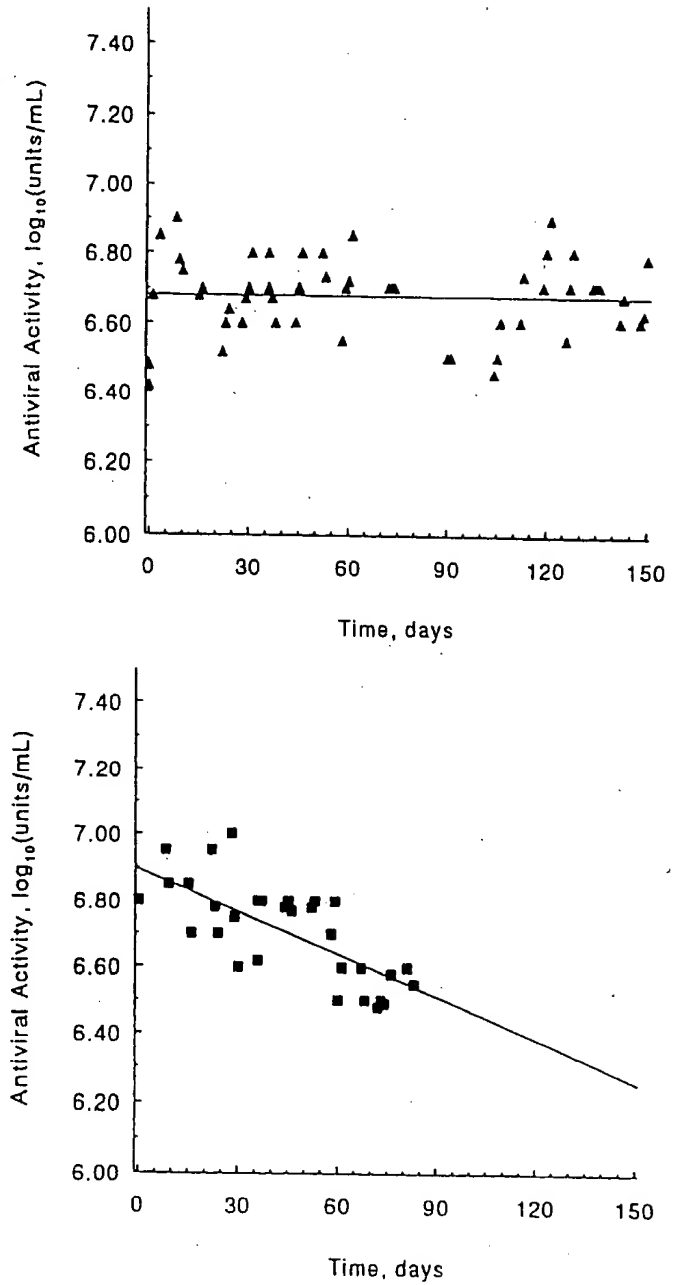


Figure 2. Stability of IFN-β_{ser17} (top) and IFN-β_{cys17} (bottom) to storage at −70°C. Interferon samples were thawed and the antiviral activities determined by a viral yield reduction bioassay at the indicated times. Each point represents the result of assays run in triplicate.

repeated twice-daily dosing and greatest antiviral efficacy of the molecule were observed under this dosing regimen.

3.2. Clinical Studies

Early clinical development of IFN-β, like other interferons, was directed toward its anticancer (Borden *et al.*, 1988, 1992; Quesada *et al.*, 1982; Reinhart, 1986) antiviral (Higgins *et al.*, 1986) and antiinfective indications (Schonfeld *et al.*, 1984). As for other indications, Jacobs *et al.* (1981) used natural IFN-β intrathecally in multiple sclerosis (MS) patients suspecting that the disease was caused by a viral infection. They reported a significant reduction in exacerbations experienced by the patients (Jacobs *et al.*, 1987). While the mechanism of action of IFN-β in MS is not fully understood, one or a combination of IFN-β activities, e.g., antiviral (Reder and Arnason, 1985), correction of deficient IFN secretion by immune cells (Neighbor and Bloom, 1979), reversal of the effects of IFN-γ (Fertsch *et al.*, 1987) and enhancement of suppressor T-cell function (Noronha *et al.*, 1990), have been implicated. A double-blind, dose-finding pilot study in subjects with relapsing-remitting MS showed that IFN-β_{ser17}, could be administered safely at a dose of 8 million IUs every other day, and demonstrated that treatment reduced the risk of exacerbations (Knobler *et al.*, 1993). A pivotal multicenter, randomized, double-blind, placebo-controlled trial of Betaseron® was conducted in 372 ambulatory patients with relapsing-remitting MS. The Betaseron® treatment caused significant reduction in exacerbation rates (compared to the placebo group), severity of exacerbations, and accumulation of magnetic resonance imaging abnormalities in the absence of serious side effects (IFNB Multiple Sclerosis Study Group, 1993; Patty *et al.*, 1993). Betaseron® (IFN-β_{ser17} or IFN-β-1b) is currently the only approved therapy in the United States for the treatment of relapsing-remitting multiple sclerosis.*

4. PHYSICOCHEMICAL CHARACTERISTICS OF IFN-β

4.1. Primary Structure

The primary structure of IFN-β_{ser17} was determined by amino acid composition, N-terminal amino acid sequencing, and peptide mapping.

*A second therapeutic, Interferon-β-1a (Avonex®, Biogen, Cambridge, MA) was approved by the FDA for the same indication in May 1996. Interferon-β-1a is a glycosylated version of the natural interferon-β.

Table I. Amino Acid Composition of Purified IFN-φ_{ser-17}

Residue	Hydrolysis time (hr) 24	48	72	Cumulative mean[a]	Predicted value from DNA sequence[b]
Asx	16.9	16.9	16.7	16.8 ± 0.5	17
Thr	7.2	7.1	6.8	7.0 ± 0.3	7
Ser	9.7	8.8	8.4	9.7 ± 0.3[c]	10
Glx	24.4	24.6	24.7	24.5 ± 0.7	24
Gly	6.3	6.4	6.4	6.3 ± 0.2	6
Ala	6.3	6.3	6.3	6.3 ± 0.2	6
Val	4.9	5.3	5.2	5.1 ± 0.2	5
Met	3.0	3.0	3.1	3.1 ± 0.2	3
Ile	10.2	10.7	10.7	10.7 ± 0.3[d]	11
Leu	24.6	24.6	24.6	24.6 ± 0.3	24
Tyr	9.9	9.8	9.9	9.9 ± 0.3	10
Phe	9.0	9.2	9.2	9.2 ± 0.4	9
Lys	10.6	10.7	11.1	10.8 ± 0.5	11
His	4.8	4.8	4.9	4.8 ± 0.2	5
Arg	11.1	10.9	11.1	11.0 ± 0.4	11
Trp	2.5	—	—	2.5 ± 0	3
Cys	2.0	—	—	2.0 ± 0.1[e]	2
Pro	1.0	—	—	1.0 ± 0.1	1

[a]The numbers representing the mean residues/molecule are averages from four separate hydrolysis series, each performed in duplicate. Cumulative mean values represent three hydrolysis times except where indicated. Uncertainties represent half the range of values averaged from cumulative mean.
[b]NH_2-terminal methionine omitted.
[c]24 hr values only.
[d]48 and 72 hr values only.
[e]Analyzed separately from the timed hydrolyses by performic acid oxidation.

4.1.1. AMINO ACID COMPOSITION

The primary amino acid sequence of IFN-β_{ser17} consists of 165 amino acids. The amino acid composition was experimentally determined to be similar to that predicted from the DNA sequence. Table I presents these data.

4.1.2. N-TERMINAL AMINO ACID SEQUENCE

Purified IFN-β_{ser17} was analyzed by N-terminal amino acid sequencing by subjecting it to automated Edman degradation in a Beckman Model 890M spinning-cup sequencer. The phenylthiohydantoin (PTH) amino acid derivatives formed in the instrument were identified using isocratic reversed-phase HPLC. These data, pre-

Table II. Partial Amino Acid Sequence of Purified IFN-β_{ser-17}

Residue number	Major residue	Yield (nmol)[a]	Minor residue	Yield (nmol)
1	Ser[b]	1.40		
2	Tyr	22.1		
3	Asn	13.9	Asp	0.91
4	Leu	17.0		
5	Leu	20.4		
6	Gly	14.8		
7	Phe	17.8		
8	Leu	15.4		
9	Gln	13.0	Glu	2.54
10	Arg	1.07		
11	Ser			
12	Ser			
13	Asn	6.46	Asp	0.51
14	Phe	9.05		
15	Gln	7.88	Glu	1.43
16	Ser			
17	Gln	6.94	Glu	1.72
18	Lys	2.31		
19	Leu	10.3		
20	Leu	10.2		
21	Trp	3.37		
22	Gln	4.54	Glu	2.11
23	Leu	9.93		
24	Asn	2.10	Asp	1.10
25	Gly	3.69		
26	Arg	1.94		
27	Leu	4.04		
28	Glu	4.37		
29	Tyr	4.17		
30	Cys[c]			

[a]A 35.7nmol sample of IFN-β_{ser-17} was subjected to automated Edman degradation, and the PTH amino acids were analyzed by reverse-phase HPLC.
[b]Serine was recovered primarily as PTH-dehydroserine which could be detected at 313 nm.
[c]Cysteine was identified as PTH-cystine. Dehydroserine and cystine were not quantitated.

sented in Table II, indicate that the first 30 amino acids from the N-terminus yielded an amino acid sequence identical to the amino acid sequence (minus the methionine residue on the N-terminus) predicted by the DNA sequence of the IFN-β_{ser17} gene. The N-terminal methionine of mature human IFN-β is used in *E. coli* as the initiation codon to direct the synthesis of the human protein. After initiation of translation, the N-terminal methionine is removed in *E. coli* by the enzyme methionine amino peptidase (MAP, Ben-Basset *et al.*, 1987). The removal of the N-terminal methionine from newly synthesized proteins by MAP is dependent on the identity of the penultimate residue and the biosynthetic rate of the recombinant protein. In the case of IFN-β_{ser17} in the production strain used for manufacturing, the removal of amino

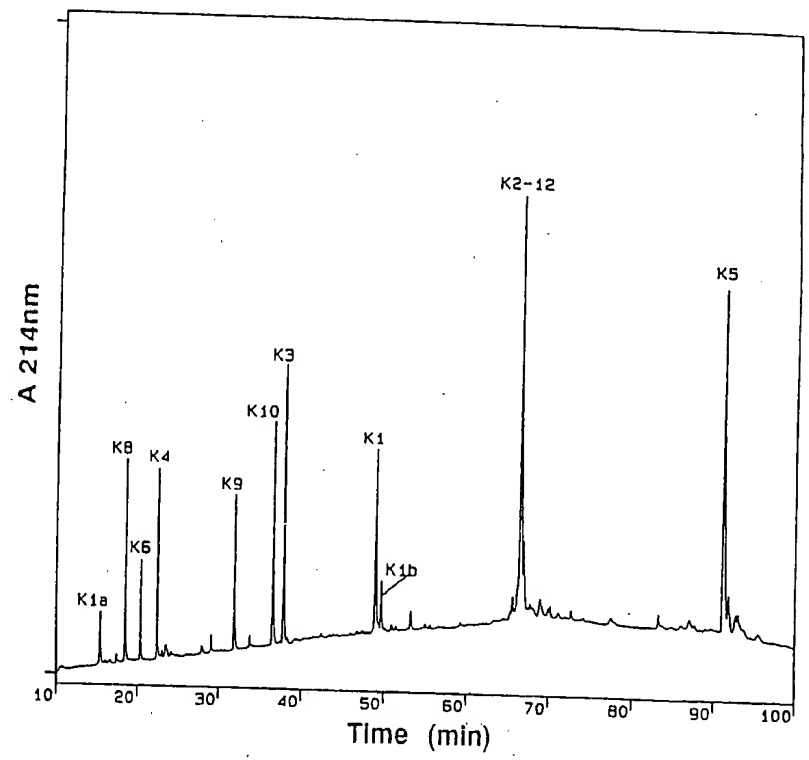


Figure 3. Peptide map of IFN-β_{ser17} digested by Lys-C. Peaks are labeled according to the order in which the corresponding peptides occur in the IFN-β_{ser17} molecule (see Fig. 4). All expected peaks are displayed, with the exception of a tripeptide (K7) and a dipeptide (K11), which elute in the unretained peak from the RP-HPLC column. Two additional peaks, "K1a" and "K1b," resulting from cleavage of Arg_{11} and Ser_{12} bond, are seen as well.

terminal methionine is very efficient, resulting in IFN-β_{ser17} with a homogeneous amino terminal albeit one residue less than that predicted by the DNA sequence.

4.1.3. PEPTIDE MAPPING

The entire amino acid sequence of IFN-β_{ser17} was determined by peptide mapping using lysyl endopeptidase Lys-C. The Lys-C peptide map in conjunction with other protein fragmentation methods provided overlapping amino acid sequences for the entire IFN-β_{ser17} molecule. The results obtained provided the entire sequence of the IFN-β_{ser17} molecule and is identical to that predicted by the DNA sequence. Figure 3 shows a typical peptide map of this molecule.

Figure 4 displays the sequence of IFN-β_{ser17} showing the cleavage sites of Lys-C. Amino acid analysis, amino acid sequence analysis, and mass spectrometry of the peptides generated by Lys-C digestion confirmed that the generated peptide fragments were identical to those predicted.

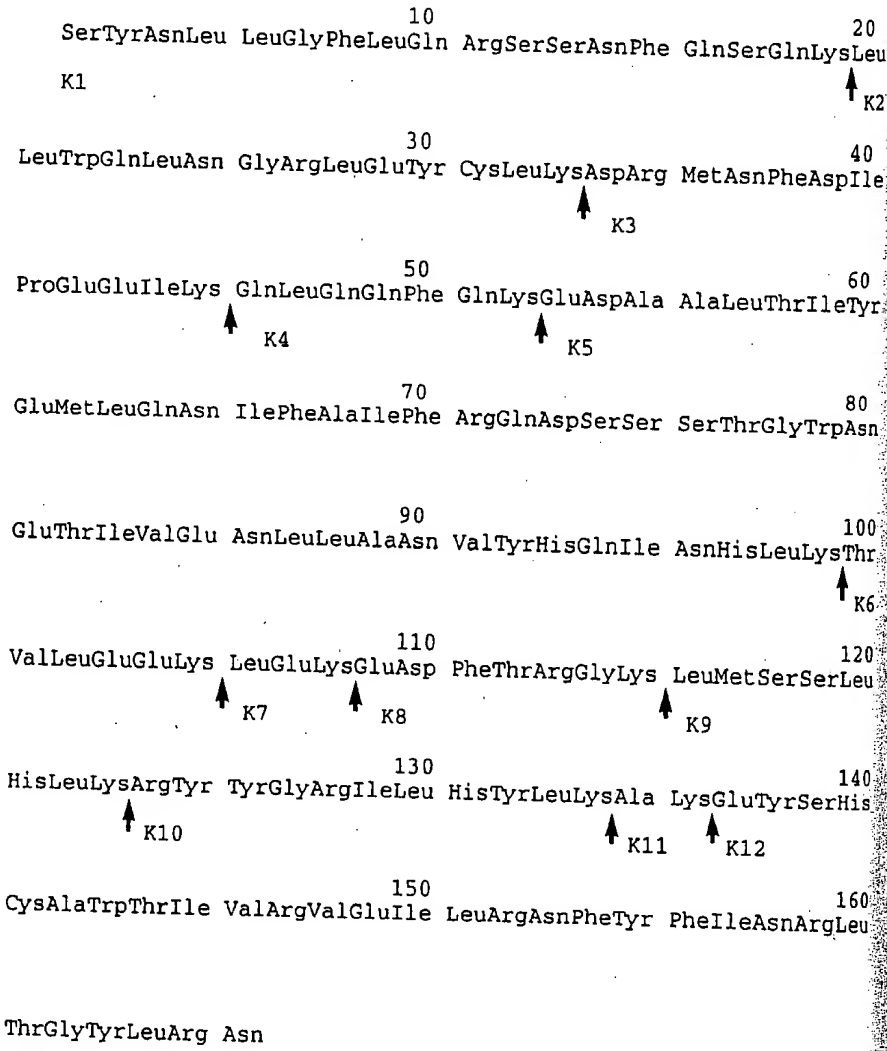


Figure 4. Amino acid sequence of IFN-β_{ser17} showing sites of proteolysis by Lys-C. Residues are numbered as in native IFN-β. Lys-C cleavage sites are indicated by bold arrows. Names of the theoretical fragments generated by Lys-C proteolysis appear beneath the sequence near their N-terminal ends.

4.2. Secondary and Tertiary Structure

4.2.1. CD AND NMR SPECTROSCOPY

Utsumi *et al.* (1986) examined the conformation of fibroblast IFN-β (glycosylated) and *E. coli*–derived IFN-β (nonglycosylated) by circular dichroism (CD) and ^{1}H nuclear magnetic resonance spectroscopy. The two interferon preparations were studied by the CD and NMR methods in an acidic pH environment (pH 4.6 to 1.6) due to good solubility and stability of IFN under these conditions. The CD spectra indicated that both IFN-βs had approximately 70% α-helix content. The data indicated that fibroblast IFN-β and *E. coli*–derived IFN-β have very similar secondary structures, thus demonstrating that the lack of glycosylation of the recombinant molecule did not alter the secondary structure of the protein. Moreover, a slow conformational change was observed below pH 2.0 which was thought to induce the disruption of β-sheets. NMR analysis were used to study the folding of IFN-β and the two spectra showed that both fibroblast and recombinant IFN-β molecules possess characteristic features of globular proteins. The NMR data also confirmed the low-β-sheet content of IFN-β.

Boublik *et al.* (1990) recently examined the relationship between the conformation and antiviral activity of *E. coli*–derived rIFN-β. The extent of ordered secondary structure was determined by CD spectroscopy in various buffer conditions. In contrast to the work of Utsumi *et al.* described above, they reported α-helical content of 40% to 50% in the pH range 2.9 to 7.2. At pH 2.9, IFN-β exhibited maximum stability to heat denaturation and highest antiviral activity. It was found that both helicity and antiviral activity of the IFN-β decrease in parallel with denaturation by urea, heat, or repeated freeze-thaw cycles. These authors also displayed the primary structure of rIFN-β in the form of a two-dimensional helical surface. Using a computer program, the potential for helix formation was calculated based on the knowledge of the primary structure. Using this model, the α-helical content of hIFN-β was estimated to be approximately 35%.

Acharya *et al.* (1985) compared the conformations of IFN-β_{cys17} and IFN-β_{ser17} by CD spectroscopy to assess whether the single amino acid substitution induces significant secondary structure changes. The studies were performed in the presence of 0.1% sodium dodecyl sulfate (SDS) at neutral pH. SDS was added to render the proteins soluble under these conditions. Both recombinant variants of IFN-β exhibited essentially the same CD spectra, consisting of approximately 35% to 40% α-helical content. These data on α-helix content of IFN-β_{ser17} and IFN-β_{cys17} were in good agreement with that obtained by Boublik *et al.* (1990) for rIFN-β and the value estimated for native human IFN-β.

The far-ultraviolet CD spectrum of IFN-β_{ser17} is presented in Fig. 5. This shows α-helical content of 45% with the rest being β-sheet.

4.2.2. FLUORESCENCE SPECTROSCOPY

Fluorescence spectroscopy has successfully been used to characterize conformational properties of IFN-α (Vincent *et al.*, 1992) and other proteins (Poklar *et al.*, 1994). The IFN-β molecule has three tryptophan residues which are located at positions 22, 79, and 143 in the sequence. The fluorescence emission maximum of IFN-β under physiological pH conditions occurs at 338 nm. In contrast, free tryptophan under identical conditions exhibits an emission maximum at 351 nm. These data indicate that the tryptophan residues within the IFN-β molecule resides in a highly hydrophobic environment (Borukhov and Strongin, 1990). Moreover, the emission

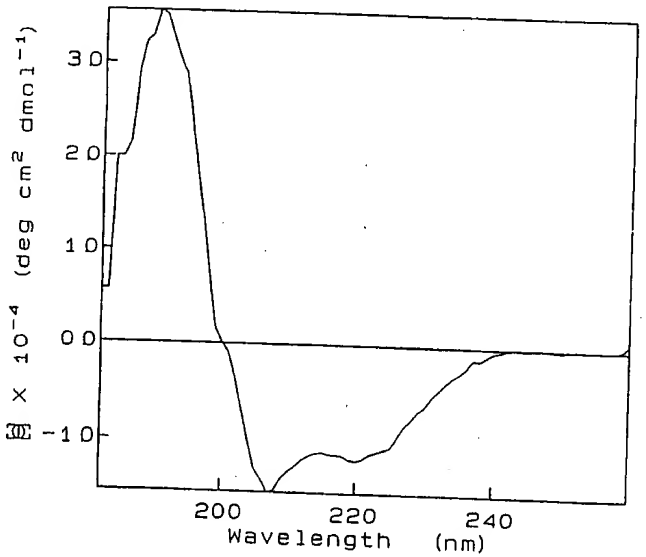


Figure 5. A representative far-ultraviolet circular dichroism spectrum of IFN-β_{ser17}.

maximum of IFN-β in its fully unfolded form (in 7 M guanidine hydrochloride) was seen at 352 nm. Further, the microenvironment of the tryptophan residues was studied in aqueous solutions at pH 2.0, 7.2, and 8.5 with KI, CsCl, and acrylamide as anionic, cationic, and neutral charge contact quenchers (Lehrer and Leavis, 1978). From these data, it was inferred that two of the three tryptophan residues of IFN-β were located near the surface of the protein. By analogy to IFN-α, tryptophan residues 22 and 143 would be expected to reside near the surface.

5. ANALYTICAL METHODS FOR EVALUATION OF PROTEIN PURITY

Besides structural information, the purity of the therapeutic protein under question is an important parameter before it is deemed suitable for use as a pharmaceutical product. The purity of the protein must also be assessed to evaluate its stability and for assignment of a shelf life to the product. Several analytical methods are used for this purpose; the primary among them being based on electrophoretic and chromatographic techniques.

5.1. SDS-PAGE

SDS-PAGE has widely been used for characterizing the purity of both native and recombinant forms of IFN-β. This method was first employed for detection of

dimers, trimers, and higher oligomers of *E. coli*–derived IFN-β_{cys17} (Colby *et al.*, 1986; Lin *et al.*, 1986; Mark *et al.*, 1984). Visualization of gels was facilitated either by staining with Coomassie Brilliant Blue dye stain or Fast Green dyes or by an anti-IFN-β monoclonal antibody after transfer on a nitrocellulose paper (Western blots). The nonreduced SDS-PAGE is capable of showing dimers, trimers, and higher oligomers. In SDS-PAGE of IFN-β samples subjected to stress by placement at high temperatures, oligomers are observed. For example, a sample of IFN-β_{ser17} (1.2 mg/ml in 50 mM sodium acetate, 10 mg SDS, 2 mM EDTA, pH 5.5) formed approximately 30% oligomers after placement at 37°C for 3 months (Geigert *et al.*, 1988). Figure 6 shows a representative densitometric scan of SDS-PAGE analysis of a reduced sample. The reduced samples exhibit only dimers and some low-molecular-weight fragments. Since the dimers are present in reduced samples, it is likely that these dimers are not linked by disulfide bonds.

5.2. Isoelectric Focusing (IEF)

The IEF method is useful for separation and visualization of charge variants of IFN-β. Utsumi *et al.* (1987) compared the IEF profiles of fibroblast IFN-β and *E. coli* produced rIFN-β on silver stained gels. Whereas fibroblast IFN-β exhibited three distinct bands with pI of 8.9 ± 0.1, 8.6 ± 0.1, and 7.8 ± 0.1, the rIFN-β showed a single, trailing band at pI of 8.9 ± 0.1. The heterogeneity in the fibroblast preparation is ascribed to the presence of varying amounts of sialic acid on the carbohydrate moiety of the molecule. All three variants possessed antiviral activities. The trailing

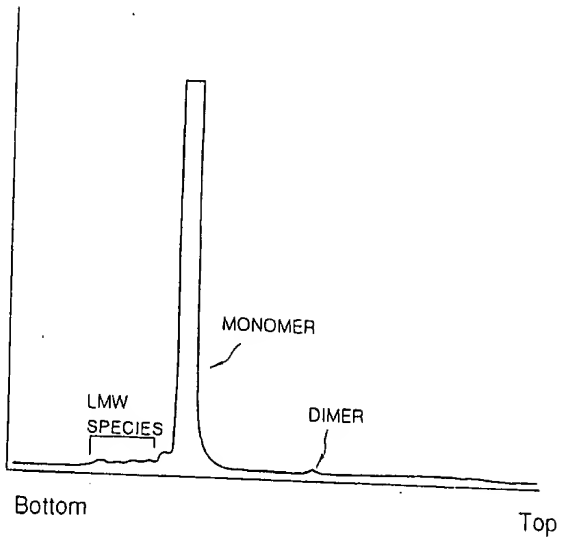


Figure 6. SDS-PAGE gel scan of reduced IFN-β_{ser17}. Sample reduced with 2-mercaptoethanol was run on a 12–15% linear gradient polyacrylamide gel, stained with Fast Green and

of the rhIFN-β band is presumably due to hydrophobic interaction between the protein and the acrylamide gel. IFN-β_{ser17} was electrofocused using the nonionic surfactant polyoxyethylene-12-lauryl ether (Laureth 12) to maintain the IFN-β_{ser17} solubility (Hershenson and Thomson, 1989). Because of the difficulties in calibrating IEF gels in the highly basic range (>pH 9), the pI for IFN-β_{ser17} was initially assigned by Hershenson as 9.6–9.7. Later, a more accurate calibration of the IEF gel was made, and a pI of 9.2 ± 0.1 was assigned.

5.3. RP-HPLC

Utsumi *et al.* (1987) reported the RP-HPLC profiles of fibroblast IFN-β and *E. coli*–derived IFN-β. They observed that the recombinant IFN-β was retained longer on the column than the fibroblast IFN-β, indicating that the former was more hydrophobic than the latter.

A representative RP-HPLC chromatogram of IFN-β_{ser17} is shown in Fig. 7. The second peak to elute (peak B) from the column represents the main IFN-β_{ser17} species. The first peak is known as peak A. Peak B can be converted to peak A under conditions specific for oxidation of methionines in proteins, suggesting that peak A is an IFN-β_{ser17} variant containing an oxidized methionine. Site-specific mutation was used to produce IFN-β_{ser17} analogues in which alanine was substituted for methionine at 36, 62, and 117 positions, respectively. Results of RP-HPLC analysis of the methionine analogues after chemical oxidation inferred that peak A contains an oxidized methionine at amino acid position 62. This was confirmed by peptide mapping of a Lys-C digest of isolated peak A.

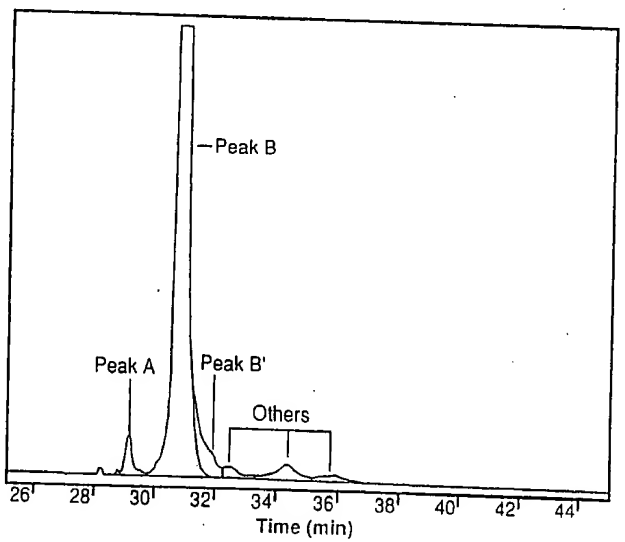


Figure 7. A RP-HPLC chromatogram of IFN-β_{ser17}. Reversed-phase high-performance liquid chromatography was conducted using a Vydac C_4 column. A gradient of 10% acetonitrile in 0.1% trifluoroacetic acid (TFA) to 100% acetonitrile in 0.1% TFA was used, and the elution was monitored by ultraviolet absorption at 214 nm.

The shoulder on the main peak, peak B′, was isolated by collecting fractions from the eluted column and thoroughly analyzed. The two isolated species had equivalent specific activities, IEF profiles, ELISA antibody responses, and peptide maps. These results indicated that peak B′ consists of a different conformational form(s) of IFN-β_{ser17} having a primary structure identical to that of peak B but resolvable by RP-HPLC.

The peaks eluting after peak B′ are mainly oligomeric forms of the IFN-β protein. These oligomers are primarily SDS-dissociable as they are not seen in the SDS-PAGE analysis (Geigert *et al.*, 1988).

6. *IN VITRO* BIOLOGICAL ACTIVITY OF IFN-β

The potency of IFN-β preparations are measured by *in vitro* biological activity assays. These assays are also important for assigning a shelf life for final commercial preparations and reference materials (Geigert *et al.*, 1988).

6.1. Antiviral Yield Reduction Assay

For measuring the antiviral activity of IFN-β, a virus yield reduction assay is employed. IFN-β containing samples are first added to GM2504 fibroblast cell for 24 hr at 37°C in a 7% CO_2 atmosphere. The cells are infected with 10^6 pfu of vesicular stomatitis virus (VSV) and incubated for 50 min. The cells are rinsed with Dulbecco's modified Eagle's medium to remove unadsorbed VSV and further incubated at 37°C for 24 hr in a 7% CO_2 atmosphere. The reduction in virus production as a result of the added IFN-β was measured in a plaquing assay by transferring the supernatant and adding to baby hamster kidney cells followed by incubation for 60 min. The number of plaques is inversely proportional to IFN-β activity. A standard curve is generated using a reference preparation of IFN-β from which the activity of an unknown IFN-β sample is determined. The potency of IFN-β_{ser17} in the yield reduction assay was found to be equivalent to that reported for native human IFN-β.

6.2. Cytopathic Effect Bioassay

A second assay that is used for measuring the potency of IFN-β preparations is based on the ability of IFN-β to inhibit viral cytopathic effects (Grossberg *et al.*,

1985). In this assay, IFN-β induced protection of A549 human lung carcinoma cells from infection with encaphlomyocarditis virus (ECV) is measured by a colorimetric method based on the ability of viable cells to reduce a dye 3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide (or MTT). Samples containing IFN-β are serially diluted and then A549 human lung carcinoma cells added. A dose-dependent antiviral state is induced in the cells by the interferon and the cells subsequently infected with ECV and IFN-β-induced cell protection measured by a spectrophotometric assay utilizing the MTT stain. The mitochondrial enzymes in viable cells reduce MTT to a dark blue formazan product which exhibits peak absorbance around 580 nm after solubilization with an alcohol/detergent solution (Mossman, 1983). Potency of IFN-β samples are determined relative to the National Institute of Health recombinant IFN-β reference material which is included on each assay plate. An interassay precision of approximately 15% has been recorded for this assay. Betaseron potency determined using the CPE assay is equivalent to potency obtained using the antiviral yield reduction assay.

7. FORMULATION STUDIES

7.1. Solubility Aspects

7.1.1. SOLUBILITY OF IFN-β_{ser17}

One major challenge with *E. coli*–derived IFN-β, partly due to it being unglycosylated, is its strongly hydrophobic character. This property of IFN-β is encountered time and again during its production and analyses. Lin *et al.*, (1986) report that IFN-β_{ser17} can be solubilized at neutral pH in the presence of surfactants such as 0.1% SDS or chaotropic agents such as 4 M guanidine hydrochloride at concentrations in the range 1–5 mg/ml. The ready solubility of IFN-β in SDS-containing solutions has been utilized throughout the purification procedure described by Lin and co-workers. Hershenson and Thomson (1989) reported the use of a nonionic surfactant (Laureth 12) for solubilizing IFN-β_{ser17} for the purpose of running an IEF gel on the protein. Utsumi *et al.* (1987) described the hydrophobicity of the *E. coli*–derived IFN-β based on longer retention of the recombinant molecule on the RP-HPLC column as compared to the retention of the fibroblast human IFN-β.

The rIFN-β_{ser17} protein is sparingly soluble (<0.05 mg/ml) at neutral pH on its own. The protein is fairly soluble (at approximately 1 mg/ml concentrations) at acidic pHs (pH 3 and below) or strongly alkaline pHs (pH 10 and above). The low solubility of this protein in the absence of stabilizers is most likely due to its hydrophobic nature. The protein tends to precipitate out due to protein–protein aggregate formation presumably through hydrophobic interactions at neutral or near-neutral pHs in the absence of solubilizing agents. These aggregates are "reversible" as they are

rendered soluble again by readdition of a solubilizer such as 0.1% SDS (Fernandes and Taforo, 1991). These data are similar to the results obtained for human fibroblast IFN-β by Utsumi *et al.* (1989). These authors reported that IFN-β formed predominantly tetrameric aggregates through hydrophobic interaction which were dissociable by 1% SDS or 1% lithium dodecyl sulfate (LDS). These tetramers were seen by size-exclusion chromatography but migrated as monomers on SDS-PAGE. Moreover, tetramers retained only 10% of the biological activity displayed by the IFN-β monomeric form but retained full activity upon 1% SDS addition.

While solubility of IFN-β in other solvent systems has not been studied in a systematic manner, selected reports present such information in an indirect way. Thus, Utsumi *et al.* (1987) used a 100 μg/ml solution of *E. coli*–derived rIFN-β in a 10 mM sodium phosphate buffer (pH 6.8) containing 0.5 M NaCl and 40% ethylene glycol. In the same report, the authors describe the use of a 2 mg/mL rIFN-β solution in 10 mM sodium phosphate buffer prepared with deuterium oxide (pD 6.8) containing 0.5 M NaCl and 40% perdeuterated ethylene glycol for NMR studies. In agreement with these data, Boublik *et al.* (1990) used 0.5 mg/ml solutions of *E. coli*–derived IFN-β in 50% ethylene glycol, 1 M NaCl, and 50 mM sodium phosphate (pH 7.2) for their studies. Solutions of 50 μg/ml rIFN-β were also prepared in 50% ethylene glycol in a citric acid–sodium phosphate buffer (pH 2.9) and ammonium acetate–NaCl buffer (pH 5.1) for CD spectral studies. Boublik *et al.* also reported that ethylene glycol had strong cryoprotective and helix-promoting effects on IFN-β and that IFN-β was fully active in these systems. These studies demonstrate that rIFN-β has reasonable solubility in 40–50% ethylene glycol perhaps in the presence of 0.5 to 1 molar NaCl. No information regarding solubility of IFN-β in glycerol, propylene glycol, and polyethylene glycol exists currently.

7.1.2. SOLUBILITY-ENHANCING STRATEGIES USED FOR IFN-β_{ser17}

rIFN-β_{ser17} is readily soluble under physiological pH conditions in the presence of the anionic surfactant SDS. Reference preparations of IFN-β_{ser17} in 0.1% SDS are described by Geigert *et al.* (1988). The minimum concentration of SDS required for solubility of 1 mg of IFN-β_{ser17} at pH 7.0 was found to be approximately 660 μg. The amount of SDS needed for IFN-β_{ser17} solubility could be reduced to 175 μg/mg of the protein by addition of 1 mg of a nonionic surfactant polysorbate-80 (Durafax-80, Durkee Chemicals). These data indicate that SDS is a more effective solubilizer for IFN-β_{ser17} than polysorbate-80. These results are in excellent agreement with data from Utsumi *et al.* (1989), who reported that SDS and LDS are effective solubilizers for rIFN-β.

A number of nonionic surfactants were evaluated for solubilization of this hydrophobic protein (Shaked *et al.*, 1993). The solubility of IFN-β_{ser17} was evaluated using an ultracentrifugation assay. In this assay, recovery of the IFN-β_{ser17} protein in the supernatant of a test solution at a given protein concentration (usually 250 to

500 μg/ml) after subjecting it to ultracentrifugation at 35,000 g for 1 hr at ambient temperature was measured. A recovery value of 80% protein in the supernatant was considered as an evidence of good solubility by this test. While this method does not provide the absolute maximum solubility of a protein in the test solution, it is useful for measuring solubility of the protein under rigorous conditions. In addition, it is a valuable tool for screening effective solubilizers for a given protein concentration and has often been used as such in the biochemical literature (Schein, 1990). A large number of nonionic surfactants were evaluated to aid solubilization of IFN-β_{ser17} (Hershenson *et al.*, 1989). Selected results from the ultracentrifugation screening are shown in Fig. 8.

Four formulations of IFN-β_{ser17}, containing surfactants such as Laureth 12 (trade name Trycol LAL12), an oxyalkylated alcohol (trade name Plurafac C-17), octoxynol-30 (trade name Triton X305), polyethylene glycol-8-oleate (trade name Nopalcol 4-O), or their mixtures were selected for further optimization studies (described next) based on the visual clarity, UV absorption, and ultracentrifugation data. A complete

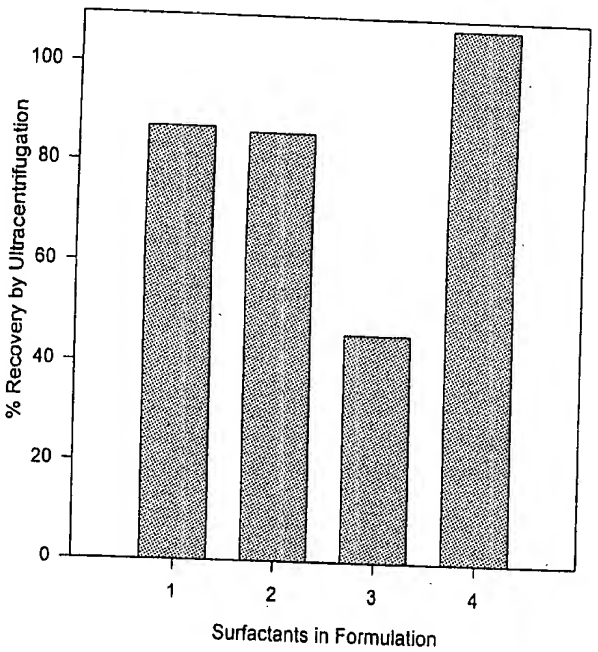


Figure 8. Comparison of four surfactant systems for formulation of IFN-β_{ser17}. Formulations containing 0.25 mg/mL IFN-β_{ser17} in 10 mM sodium phosphate and one of the following surfactant(s): 0.15% laureth-12 (1), 0.10% oxyalkylated alcohol (Plurafac C-17), (2) a combination of 0.10% octoxynol-30 and 0.05% PEG-8-oleate (3) or a combination of 0.10% laureth-12 and 0.05% PEG-8-oleate (4) were evaluated by the ultracentrifugation assay. Individual bars show the recovery of IFN-β_{ser17} in the top half of the solution after centrifugation at 35,000 g for 1 hr by A_{280} measurements.

cross-reference of generic and trade names of these surfactants is available (Ash and Ash, 1993). A comparison of buffers indicated that for lyophilized IFN-β_{ser17} formulations, sodium phosphate was better for maintaining solubility of the protein upon reconstitution than sodium citrate and sodium maleate buffers. It was also surmised by Hershenson *et al.* (1989) that pH change caused by the well-known crystallization of the disodium phosphate component of the phosphate buffer during freezing may have helped in preserving the solubility of IFN-β_{ser17}.

For maintaining solubility of IFN-β_{ser17} after lyophilization, formulations with potential bulking agents were evaluated by the ultracentrifugation assay. The data, shown in Fig. 9, indicate that dextrose or a combination of dextrose and mannitol were suitable for this purpose while dextran, mannitol, or a dextrose/glycine mixture were unable to preserve solubility of IFN-β_{ser17} upon reconstitution (Hershenson *et al.*, 1989).

Finally, carrier proteins, such as human serum albumin (HSA) and plasma protein fraction (PPF), have also been found to be useful for rendering the IFN-β_{ser17}

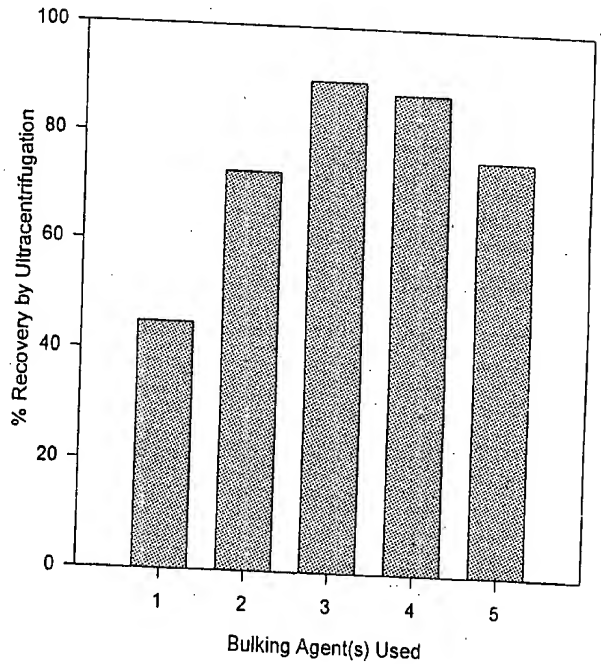


Figure 9. Effect of different bulking agents on the solubility of IFN-β_{ser17} upon reconstitution of the freeze-dried product. Formulations contained IFN-β_{ser17} (0.25 mg/ml) in 0.15% laureth-12 and 10 mM sodium phosphate buffer (pH 7) and one of the following bulking agents: 2.0% dextran (1), 2.0% mannitol (2), 2.0% dextrose (3), a combination of 0.1% dextrose and 2.0% mannitol (4) or a combination of 0.1% dextrose and 2.0% glycine (5). Individual bars show the recovery of IFN-β_{ser17} in the top half of the solution after ultracentrifugation at 35,000 g for 1 hr.

soluble under physiological pH conditions (Fernandes and Taforo, 1991). IFN-β_{ser17} could be solubilized by adding HSA to a 1:50 weight/weight (wt/wt) ratio. Formulations at 1 mg/ml IFN-β_{ser17} concentration were prepared using the 1:50 IFB-β_{ser17} ratio. PPF, which consists of 83% HSA and a maximum of 17% globulins (α- and β-), was also shown to solubilize IFN-β_{ser17} at similar wt/wt ratios. Solubilization of IFN-β_{ser17} in HSA and PPF solutions is thought to occur via interaction between the hydrophobic segments of IFN-β_{ser17} and HSA.

7.2. Parenteral Formulations of IFN-β_{ser17}

A recombinant form of IFN-β, interferon-β-1b or IFN-β_{ser17} (Betaseron®, a product of Chiron Corporation), is available commercially in the United States since 1993. Betaseron® is supplied as a lyophilized powder consisting of 0.25 mg of interferon-β-1b and contains 12.5 mg each of human serum albumin and dextrose. Appropriate amounts of sodium hydroxide and hydrochloric acid may have been used for adjustment of pH of the solution to 7.5. A diluent vial containing 0.54% sodium chloride is supplied along with Betaseron®. This concentration of sodium chloride yields an isotonic solution upon reconstitution of lyophilized Betaseron® as directed in the package insert. Each vial of Betaseron® is reconstituted with 1.2 ml of the supplied diluent and 1.0 ml of the reconstituted solution is injected subcutaneously by patients for the treatment of relapsing-remitting multiple sclerosis (Betaseron, Physician Desk Reference, 1995).

7.3. Long-Acting Formulations of IFN-β_{ser17}

Considerable research has been done to prolong the *in vivo* delivery of IFN-β_{ser17}. To enhance solubility and *in vivo* half-life of the recombinant molecule, it was modified by attachment of water-soluble polymers such as polyethylene glycol (PEG) and polyoxyethylene glycol (POG) (Katre and Knauf, 1990). Attachment with such polymers has successfully been used for altering the hydrodynamic radius of the resulting PEG–protein yielding a product with a desired *in vivo* half-life (Knauf *et al.*, 1988). The solubility of IFN-β_{ser17} could be greatly enhanced by PEG-attachment while maintaining the bioactivity of IFN-β. Similarly, the *in vivo* half-life of IFN-β_{ser17} was enhanced severalfold by the modification (Katre and Knauf, 1990).

Liposomal formulations of IFN-β have also been evaluated. Felgner and Epstein (1985) described a liposomal formulation of IFN-β_{ser17} made by hydrating a lyophilized mixture of multilamellar vesicles with an IFN-β_{ser17} solution. The encapsulated IFN-β retained full antiviral activity. The controlled release of IFN-β_{ser17} from this system was demonstrated in a mouse model after intramuscular injection. In control animals, free IFN-β_{ser17} disappeared from the injection site in 1 day while IFN-β_{ser17} from liposomes was maintained at the injection site up to 9 days. In a subsequent study, this formulation was tested in a Simian *Varicella* virus infected African green monkey model (Eppstein *et al.*, 1989). It was observed that intramuscularly injected liposomal IFN-β_{ser17} resulted in a sustained release of the IFN-β from the injection site. Finally, the liposomal preparation exerted antiviral efficacy in the primate model superior to that obtained with the identical dosing regimen of free IFN-β_{ser17}.

The biodegradable polylactide-*co*-glycolide (PLG) polymer system has also been used for the controlled release of rIFN-β_{ser17} (Eppstein and Schryver, 1990). The protein was incorporated in the PLG matrix by a spray-casting technique. Prior to the encapsulation process, the IFN-β_{ser17} was spiked with a small amount of radiolabeled (^{125}I-IFN-β_{ser17}). No loss in the antiviral activity of IFN-β_{ser17} was seen by the process of encapsulation. Hollow cylindrical devices of PLG containing IFN-β_{ser17} films (300 μ thick, 5 mm long with ~0.5 mm external diameter) were sterilized by gamma irradiation and implanted subcutaneously in mice. No information on the effect of gamma irradiation on the integrity of encapsulated protein was provided in the report. The devices were removed surgically at periodic intervals and assayed for remaining radioactivity. Release of IFN-β_{ser17} was extended over a period of approximately 70 days.

8. STABILITY OF IFN-β

8.1. Stability-Indicating Assays

Several stability-indicating methodologies for IFN-β are available. The choice of the method depends upon the nature of the formulation. In formulations containing a carrier protein such as albumin, the normal methods used for the protein-purity analyses of IFN-β can be difficult because of interference from the carrier protein. In such cases, methods based on immunological detection of IFN-β are employed. Thus, enzyme-linked immunosorbent assays (ELISAs) based on monoclonal antibodies raised against the rhIFN-β molecule are used for quantification of IFN-β in the presence of a carrier protein. Similarly, the SDS-PAGE gels used for evaluation of oligomers and fragments of the IFN-β protein, are visualized by monoclonal antibodies after transfer to a nitrocellulose paper in the Western blot format. A common limitation of the immunological methods is that they can only detect only certain epitopes on the molecule.

In formulations utilizing no carrier protein, the regular SDS-PAGE method has been applied for detection and quantitation of oligomers and fragments of the IFN-β_{ser17} protein (Geigert *et al.*, 1988). Additionally, the RP-HPLC method has been used which is capable of tracking increases in the oxidized methionine form as well as oligomers of IFN-β. Based on RP-HPLC data of IFN-β_{ser17} formulated in the absence of a carrier protein, no increase in the oxidized methionine IFN-β_{ser17} peak

was observed even after placement at 37°C for 3 months. By RP-HPLC, only oligomer formation was observed in the IFN-β_{ser17} product. These oligomers were not seen by the SDS-PAGE method, indicating that the oligomers were SDS-dissociable.

8.2. Stability of IFN-β_{ser17}

As expected, stability of IFN-β_{ser17} is a function of the formulation parameters. In the noncarrier protein solution formulation of IFN-β_{ser17} (per milliliter composition: 1.2 mg IFN-β_{ser17}, 10 mg SDS in 50 mM sodium acetate and 2 mM EDTA, pH 5.5) described by Geigert *et al.* (1988), an Arrhenius fit of the data was attempted. Based on the SDS-PAGE and RP-HPLC data, a t_{90} (i.e., time to reach 90% IFN purity) of 7 years was predicted at 5°C (2–8°C). An activation energy of 24 kcal/mole was reported for the rate of IFN-β degradation.

In IFN-β_{ser17} formulations containing human serum albumin as a solubilizing and stabilizing agent, the biological potency of IFN- β_{ser17} was reported during storage of the lyophilized product at 5°C (Geigert *et al.*, 1987). While no changes in the potency of the three subject formulations were observed at 5°C over 2 years, temperature-dependent decreases in this parameter were observed at elevated temperatures (25, 37, 55, 75, and 80°C). Based on the elevated temperature data, an activation energy of 25 kcal/mole was obtained.

Figure 10 presents data on the stability of the Betaseron® product as measured by its biological potency.

IFN-β_{ser17} formulations have also been evaluated by linear nonisothermal stability (LNS) studies (Geigert *et al.*, 1987; Jameson *et al.*, 1979). In this method, lyophilized IFN-β_{ser17} formulations were heated from 50°C to 80°C at a linear rate of 1.5°C/hr and samples were withdrawn at pre-determined set points and analyzed for biological potency. This method is best used for comparing different formulations within a short time frame. For example, Geigert *et al.* (1987) evaluated three slightly different IFN-β_{ser17} formulations based on HSA. The three formulations contained 0.06, 0.30, and 1.20 mg of IFN-β_{ser17} with 15, 15, and 60 mg of HSA, respectively; and each formulation used 15 mg dextrose as a bulking agent. All formulations contained ≤1% moisture by weight at the start of the study. Formulation containing the highest amount of HSA showed maximal stability by the real-time, multiple isothermal, and LNS studies demonstrating the usefulness of this technique for comparative purposes.

Lyophilized formulations of IFN-β_{ser17} based on the surfactant Laureth 12 were analyzed by the LNS method (Fig. 11). An HSA-based formulation was used as a control in this study as a relative relationship of stability indicated by the real-time and LNS studies had already been established for this formulation. IFN-β_{ser17} formulations containing Laureth 12 with either dextrose or dextrose/mannitol appear to have potency stability characteristics similar to that of the HSA formulation of IFN-β_{ser17}.

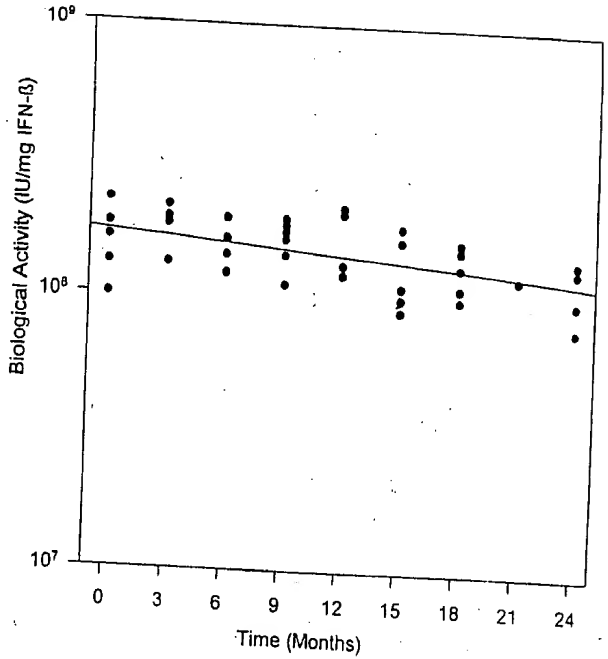


Figure 10. Potency stability of Betaseron® at 5°C. Stability of six different batches of IFN-β_{ser17} formulated with human serum albumin as measured by the virus yield reduction biological activity assay as a function of the time of incubation under refrigeration conditions. The potency in international units per milligram of the IFN-β protein on a logarithmic scale on the *y*-axis and the time of incubation at 5°C on the *x*-axis are shown.

9. CONCLUSIONS

In this chapter, we have attempted to provide a brief historical perspective on the development of human recombinant interferon beta, with a special emphasis on the research and development of Betaseron®, a recombinant human IFN-β, as a therapeutic protein drug. Brief summaries of the molecular biology and protein chemistry of IFN-β and its preclinical and clinical evaluations are presented to familiarize the reader with the complexity of the drug development process as it applies to therapeutic protein molecules.

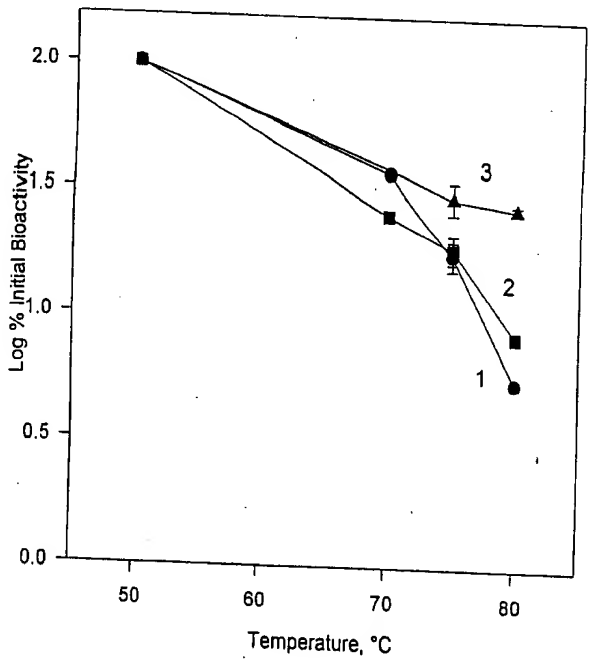


Figure 11. Stability comparison of three different formulations of IFN-β_{ser17} by linear nonisothermal studies. IFN-β_{ser17} formulations (0.25 mg/ml) containing 1.25% HSA and 1.25% dextrose (1), 0.15% laureth-12 and 5% dextrose (2), or 0.15% laureth-12 and 5% mannitol (3) are compared. The *x*-axis represents the temperature at which the IFN-β sample was withdrawn during linear nonisothermal heating and the *y*-axis shows the biological potency of the sample measured by the yield reduction bioassay and represented as the logarithmic of the initial value.

Betaseron® was one of the first few recombinant protein drugs to be tested in human clinical trials at the time when the recombinant DNA technology was at its infancy. We have described to the reader some of the difficulties that were encountered during its development, especially due to the strong hydrophobic nature of the molecule. We have presented the important physicochemical properties of this protein and a description of the analytical methods used for defining its purity. Finally, IFN-β formulations and their stability have been discussed.

ACKNOWLEDGMENTS. The authors would like to thank the many people from Chiron and Berlex who have worked on development of Betaseron® from its inception to the present time. This important therapy would not have been available to multiple sclerosis patients without their hard work and dedication.

REFERENCES

Acharya, A. S., Manjula, B. N., and Drummond, R. J., 1985, Conformational analysis of human fibroblast interferon cloned and expressed in *E. coli* poster presented at FASEB Meeting, Anaheim, CA.

Arabje, Y. M., Bittner, G., Yingling, J. M., Storer, B., and Schiller, J. H., 1993, Antiproliferative effects of interferons -α and -β in combination with 5-fluorouracil, cisplatin and *cis*- and *trans*-retinoic acid in three human lung carcinoma cell lines, *J. Interferon Res.* **13**:25–32.

Ash, M., and Ash, I. 1993, *Handbook of Industrial Surfactants*, Gower, Hants, UK.

Ben-Bassat, A., Bauer, K., Chang, S.-Y., Myabo, K., Boosman, A., and Chang, S., 1987, Processing of the initiation methionine from proteins: Properties of the *Escherichia coli* methionine aminopeptidase and its gene structure, *J. Bacteriol.* **169**:751–757.

Betaseron, 1995, *Physician Desk Reference*, pp. 622–626.

Borden, E. C., Hawkins, M. J., Sielaff, K. M., Storer, B. M., Schiesel, J. D., and Smalley, R. V., 1988, Clinical and biological effects of recombinant interferon-β administered intravenously daily in phase I trial, *J. Interferon Res.* **8**:357–366.

Borden, E. C., Kim, K., Ryan, L., Blum, R. H., Shiraki, M., Tormey, D. C., Comis, R. L., Hahn, R. G., and Parkinson, D. R., 1992, Phase II trials of interferons-α and -β in advanced sarcomas, *J. Interferon Res.* **12**:455–458.

Borukhov, S. I., and Strongin, A. Y., 1990, The intrinsic fluorescence of the recombinant human leukocyte interferon-αA and fibroblast-β1, *Biochem. Biophys. Res. Comm.* **169**:282–288.

Boublik, M., Moschera, J. A., Wei, C., and Kung, H., 1990, Conformation and activity of recombinant human fibroblast interferon-β, *J. Interferon Res.* **10**:213–219.

Chiang, J., Gloff, C. A., Soike, K. F., and Williams, G., 1993, Pharmacokinetics and antiviral activity of recombinant human IFN-β_{ser17} in African green monkeys, *J. Interferon Res.* **13**:111–120.

Colby, C. B., Geigert, J. H., Ruzicka, F. J., Schiller, J. H., Willson, J. K. V., Chen, B. P., Sondel, P. M., and Borden, E. C., 1986, *In vitro* biological assessment of an interferon beta with a site specific amino acid substitution, in: *The Biology of the Interferon System* (W. E. Stewert II and H. Schellekens, eds.), Elsevier, pp. 273–278.

Day, C., Schwartz, B., Li, B.-L., and Pestka, S., 1992, Engineered disulfide bond greatly increases specific activity of recombinant murine interferon-β. *J. Interferon Res.* **12**:139–143.

Derynick, R., Remaut, E., Saman, E., Stranssens, P., De Clercq, E., Content, J., and Fiers, W., 1980, Expression of human fibroblast interferon gene in *Escherichia coli*, *Nature (London)* **287**:193–197.

Dianzani, F., and Dolei, A., 1984, From Isaacs to *Escherichia coli*, in: *Contributions to Oncology* (L. Borecky and V. Lackovic, eds.), Karger, Basel, pp. 1–14.

Eppstein, D. A., and Schryver, B. B., 1990, Controlled release of macromolecular polypeptides, U.S. Patent No. 4,962,091.

Eppstein, D. A., Van Der Pas, M. A., Gloff, C. A., and Soike, K. F., 1989, Liposomal interferon-β: sustained release treatment of simian varicella virus infection in monkeys, *J. Infect. Dis.* **159**: 616–620.

Faltynek, C. R., and Baglioni, C., 1984, Interferon is a polypeptide hormone, *Microbiol. Sci.* **1**:81–85.

Felgner, P. L., and Eppstein, D. A., 1985, Stable liposomes with aqueous-soluble medicaments and methods for their preparation. European Patent Publication No. 0172007A2.

Fernandes, P. M., and Taforo, T., 1991, Formulations for lipophilic IL-2 proteins, U.S. Patent, No. 4,992,271.

Fertsch, D., Schoenberg, D. R., German, R. N., Tou, J. Y. L., and Vogel, S. N., 1987, Induction of macrophage Ia antigen expression by rIFN-Γ and down-regulation by IFNα/β and dexamethasone are mediated by changes in steady-state levels of Ia mRNA, *J. Immunol.* **139**:244–249.

Geigert, J., Panschar, B. M., Fong, S., Huston, H., Wong, D. A., Wong, D. Y., Taforo, C., and Pemberton, M., 1988, The long-term stability of recombinant (serine-17) human interferon-β, *J. Interferon Res.* **8**:539–547.

Geigert, J., Ziegler, D. L., Panschar, B. M., Creasey, A. A., and Vitt, C. R., 1987, Potency stability of recombinant (serine-17) human interferon-β, *J. Interferon Res.* **7**:203–211.

Grossberg, S. E., Taylor, J. L., Seibenlist, R. E., and Jameson, P., 1985, Biological and immunological assays of human interferons, in: *Manual of Clinical Immunology* (N. R. Rose, H. Friedman, and J. L. Fahey, eds.), American Society of Microbiology, pp. 295–299.

Hershenson, S., Stewart, T., Carrol, C., and Shaked, Z., 1989, Formulation of recombinant interferon-β using laureth-12, a novel nonionic surfactant, in: *Therapeutic Peptides and Proteins: Formulation, Delivery and Targeting* (D. Marshak and D. Liu, eds.), Cold Spring Harbor, NY, pp. 31–36.

Hershenson, S., and Thomson, J., 1989, Isoelectric focusing of recombinant interferon-β, *Appl. Theor. Electrophor.* **1**:123–125.

Higgins, P. G., Al-Nakib, W., Willman, J., and Tyrell, D. A. J., 1986, Interferon-β_{ser} as prophylaxis against experimental rhinovirus infection in volunteers, *J. Interferon Res.* **6**:153–159.

IFNβ Multiple Sclerosis Study Group, 1993, Interferon beta-1b is effective in relapsing-remitting multiple sclerosis. I. Clinical results of a multicenter, randomized, double-blind, placebo-controlled trial, *Neurology* **43**:655–661.

Jacobs, L., O'Malley, J., Freeman, A., and Ekes, R., 1981, Intrathecal interferon reduces exacerbations of multiple sclerosis, *Science* **214**:1026–1028.

Jacobs, L., Salazar, A. M., Herndon, R., *et al.*, 1987, Intrathecally administered natural human fibroblast interferon reduces exacerbations of multiple sclerosis: results of a multicenter, double-blinded study, *Arch. Neurol.* **44**:589–595.

Jameson, P., Grieff, D., and Grossberg, S. E., 1979, Thermal stability of freeze-dried mammalian interferons. Analysis of freeze-drying conditions and accelerated storage tests for murine interferon, *Cryobiology* **16**:301–314.

Katre, N., and Knauf, M. J., 1990, Solubilization of immunotoxins for pharmaceutical compositions using polymer conjugation, U.S. Patent No. 4,917,888.

Kerr, I. M., and Stark, G. R., 1992, The antiviral effects of the interferons and their inhibition, *J. Interferon Res.* **12**:237–240.

Kirchner, H., 1986, The interferon system as an integral part of the defense system against infections, *Antiviral Res.* **6**:1–17.

Knauf, M. J., Bell, D. P., Hirtzer, P., Luo, Z. P., Young, J. D., and Katre, N. V., 1988, Relationship of effective molecular size on systemic clearance in rats of recombinant interleukin-2 chemically modified with water soluble polymers, *J. Biol. Chem.* **263**:1564–1570.

Knobler, R. L., Greenstein, J. I., Johnson, K. P., *et al.*, 1993, Systemic recombinant human interferon-β treatment of relapsing-remitting multiple sclerosis: pilot study analysis and six-year follow-up, *J. Interferon Res.* **13**:333–340.

Lehrer, S. S., and Leavis, P. C., 1978, Solute quenching of protein fluorescence, in: *Methods in Enzymology*, Vol. 49, Academic, New York, pp. 222–254.

Lin, L. S., Yamamoto, R., and Drummond, R. J., 1986, Purification of recombinant human interferon-β expressed in *Escherichia coli*, in: *Methods in Enzymology*, Vol. 119, Academic, New York, pp. 183–192.

Mark, D. F., Lu, S. D., Creasey, A. A., Yamamoto, R., and Lin, L. S., 1984, Site-specific mutagenesis of the human fibroblast interferon gene, *Proc. Natl. Acad. Sci. USA* **81**:5662–5666.

Mossman, T., 1983, Rapid colorimetric assay for cellular growth and survival: application to proliferation and cytotoxic assays, *J. Immunol. Methods* **65**:55–63.

Murray, H. W., 1992. The interferons: macrophage activation, and host defence against nonviral pathogens, *J. Interferon Res.* **12**:319–322.

Neighbor, P. A., and Bloom, B. R., 1979, Absence of viral-induced lymphocyte suppression and interferon production in multiple sclerosis, *Proc. Natl. Acad. Sci. USA* **76**:476–480.

Noronha, A., Toscas, A., and Jensen, M. A., 1990, Interferon beta augments suppressor cell function in multiple sclerosis, *Ann. Neurol.* **27**:207–220.

Patty, D. W., Li, D. K. B., *et al.*, 1993, Interferon beta-1b is effective in relapsing-remitting multiple sclerosis. II. MRI analysis results of a multicenter, randomized, double-blind, placebo-controlled trial, *Neurology* **43**:662–667.

Pestka, S., 1983, The human interferons: from protein purification and sequence to cloning and expression in bacteria: before, between and beyond, *Arch. Biochem. Biophys.* **221**:1–37.

Poklar, N., Vesnaver, G., and Lapanje, S. 1994, Denaturation behavior of α-chymotrypsinogen A in urea and alkylurea solutions: fluorescence studies, *J. Protein Chem.* **13**:323–331.

Quesada, J. R., Gutterman, J. U., and Hersh, E. M., 1982, Clinical and immunological study of beta interferon by intramuscular route in patients with metastatic breast cancer, *J. Interferon Res.* **2**: 593–599.

Reder, A. T., and Arnason, B. G. W., 1985, Immunology of multiple sclerosis, in: *Handbook of Clinical Neurology. Demyelinating Diseases*, Vol. 3 (J. C. Koetsier, G. W. Vinken, G. W. Bruyn, and H. L. Klawans, eds.), Elsevier, Amsterdam, pp. 337–395.

Reinhart, J., Malspies, L., Young, D., and Neidhart, J., 1986, Phase I/II trial of human recombinant β-interferon serine in patients with renal cell carcinoma, *Cancer Res.* **46**:5364.

Reiter, Z., 1993. Interferon—a major regulator of natural killer cell-mediated cytotoxicity, *J. Interferon Res.* **13**:247–257.

Schein, C. H., 1990. Solubility as a function of protein structure and solvent components, *Bio/Technology* **8**:308–315.

Schonfeld, A., Nitke, S., Schatner, A., Wallach, D., Crespi, M., Hahn, T., Lavavi, H., Yarden, O., Shoham, J., Doerner, T., and Revel, M., 1984, Intramuscular human interferon-β injections in treatment of condylomata acuminata, *Lancet* 1038–1042.

Shaked, Z., Stewart, T., Hershenson, S., Thomson, J. W., and Thomson, J., 1993, Formulation processes for pharmaceutical compositions of recombinant beta-interferon, U.S. Patent No. 5,183,746.

Shepard, H. M., Leung, D., Stebbing, N., and Goddel, D. V., 1981, A single amino acid change in IFN-β_1 abolishes its antiviral activity, *Nature (London)* **294**:563–565.

Soike, K. F., Chou, T.-C., Fox, J. J., Watanabe, K. A., and Gloff, C. A., 1990, Inhibition of simian *Varicella* virus infection of monkeys by 1-(2-deoxy-2 fluoro-1-D-arabinofuranosyl)-5-ethyl uracil (FEAU) and synergistic effects of combination with human recombinant interferon-β, *Antiviral Res.* **13**:165.

Soike, K. F., Eppstein, D., Gloff, C. A., Cantrell, C., Chou, T.-C., and Gerone, P. J., 1987, Effect of 9-(1,3-dihydroxy-2-propoxymethyl)guanine and recombinant human interferon alone and in combination on Simian *Varicella* virus infection in monkeys, *J. Infect. Dis.* **156**:607–614.

Utsumi, J., Yamazaki, S., Hosoi, K., Kimura, S., Hanada, K., Shimazu, T., and Shimizu, T., 1987, Characterization of *E. coli*–derived recombinant human interferon-β as compared with fibroblast human interferon-β, *J. Biochem.* **101**:1199–1208.

Utsumi, J., Yamazaki, S., Kawagucki, K., Kimura, S., and Shimizu, H., 1989, Stability of human interferon-β1: oligomeric human interferon-β1 is inactive but is reactivated by monomerization, *Biochim. Biophys. Acta* **998**:167–172.

Utsumi, J., Yamazaki, S., Hosoi, K., Shimizu, H., Kawaguchi, K., and Inagaki, F., 1986, Conformations of fibroblast and *E. coli*–derived recombinant human interferon-βs as studied by nuclear magnetic resonance and circular dichroism, *J. Biochem.* **99**:1533–1535.

Vincent, M., Sierra, I. M., Berbaron-Santos, M. N., Diaz, A., Diaz, M., Padron, G., and Gallay, J. 1992, Time-resolved fluorescence study of human recombinant interferon α_2 association state of the protein, spatial proximity of the two tryptophan residues, *Eur. J. Biochem.* **210**:953–961.

Wetzel, R., Johnson, P. D., and Czarniecki, C. W., 1983, Roles of the disulphide bonds in a human alpha interferon, in: *The Biology of the Interferon System* (E. De Maeyer and H. Schellekens, eds.), Elsevier, Amsterdam, pp. 101–120.

Zoon, K., 1987, Human interferons: structure and function, in: *Interferon 9* (I. Gresser, ed.), Academic, New York, pp. 1–13.

Series Editor: Ronald T. Borchardt
The University of Kansas
Lawrence, Kansas

Recent volumes in this series:

Volume 2 STABILITY OF PROTEIN PHARMACEUTICALS, Part A: Chemical and Physical Pathways of Protein Degradation
Edited by Tim J. Ahern and Mark C. Manning

Volume 3 STABILITY OF PROTEIN PHARMACEUTICALS, Part B: *In Vivo* Pathways of Degradation and Strategies for Protein Stabilization
Edited by Tim J. Ahern and Mark C. Manning

Volume 4 BIOLOGICAL BARRIERS TO PROTEIN DELIVERY
Edited by Kenneth L. Audus and Thomas J. Raub

Volume 5 STABILITY AND CHARACTERIZATION OF PROTEIN AND PEPTIDE DRUGS: Case Histories
Edited by Y. John Wang and Rodney Pearlman

Volume 6 VACCINE DESIGN: The Subunit and Adjuvant Approach
Edited by Michael F. Powell and Mark J. Newman

Volume 7 PHYSICAL METHODS TO CHARACTERIZE PHARMACEUTICAL PROTEINS
Edited by James N. Herron, Wim Jiskoot, and Daan J. A. Crommelin

Volume 8 MODELS FOR ASSESSING DRUG ABSORPTION AND METABOLISM
Edited by Ronald T. Borchardt, Philip L. Smith, and Glynn Wilson

Volume 9 FORMULATION, CHARACTERIZATION, AND STABILITY OF PROTEIN DRUGS: Case Histories
Edited by Rodney Pearlman and Y. John Wang

Formulation, Characterization, and Stability of Protein Drugs

Case Histories

Edited by

Rodney Pearlman
Megabios Corporation
Burlingame, California

and

Y. John Wang
Scios Nova, Inc.
Mountain View, California

Plenum Press • New York and London

Contents

1

A Compendium and Hydropathy/Flexibility Analysis of Common Reactive Sites in Proteins: Reactivity at Asn, Asp, Gln, and Met Motifs in Neutral pH Solution

Michael F. Powell

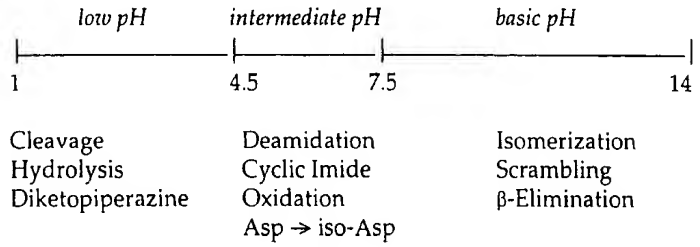
with Godfrey Amphlett, Jerry Cacia, William Callahan, Eleanor Cannova-Davis, Byeong Chang, Jeffrey L. Cleland, Todd Darrington, Linda DeYoung, Bhim Dhingra, Rich Everett, Linda Foster, John Frenz, Anne Garcia, David Giltinan, Gerry Gitlin, Wayne Gombotz, Michael Hageman, Reed Harris, Debra Heller, Alan Herman, Susan Hershenson, Maninder Hora, Rebecca Ingram, Susan Janes, Madav Kamat, Dan Kroon, Rodney G. Keck, Ed Luedke, Leonard Maneri, Carl March, Louise McCrossin, Tue Nguyen, Suman Patel, Hong Qi, Michael Rohde, Barry Rosenblatt, Nancy Sahakian, Zahra Shahrokh, Steve Shire, Cynthia Stevenson, Kenneth Stoney, Suzanne Thompson, Glen Tolman, David Volkin, Y. John Wang, Nicholas Warne, Colin Watanabe

Michael F. Powell • Department of Pharmaceutical Research and Development, Genentech, Inc., South San Francisco, California 94080.

Formulation, Characterization, and Stability of Protein Drugs, Rodney Pearlman and Y. John Wang, eds., Plenum Press, New York, 1996.

1. INTRODUCTION

The accurate prediction of protein stability under pharmaceutical formulation conditions is one of the more challenging goals in protein formulation. Almost all protein and peptide liquid formulations are designed to be at or near the pH of maximum stability of the protein, usually between pH 4.5 and 7.5. The reactions that most proteins undergo within this pH range are also narrowly defined; there are several reactions that may occur at high or low pH, but are negligible in the pH 4.5–7.5 range. Within this window of "neutral" pH, the major degradation reactions are deamidation, cyclic imide formation, iso-Asp formation, and oxidation. Other chemical reactions, including backbone cleavage (such as at the reactive Asp-Pro site), racemization, pyroglutamic acid formation, diketopiperazine formation, disulfide exchange, and others, occur predominantly at high or low pHs (Scheme 1).



Scheme 1

The goal of this compilation on the chemical reactivity of proteins is to establish boundaries for the reactivity of Asn, Asp, Gln, and possibly Met, in the context of neighboring amino acid sequence, hydrophobicity and backbone flexibility. Given a particular primary amino acid sequence, is it possible to predict with some certainty the likelihood of a particular deamidation or oxidation reaction under conditions of a liquid pharmaceutical formulation? To answer this question, we surveyed the literature for protein degradation under "typical" formulation conditions (aqueous solution, pH 4.5–7.5, 2–37°C). Our goal was to address several questions:

1. What are the predominant site(s) of chemical degradation, either deamidation or oxidation, in the proteins reported so far? Are there many exceptions to the rules already in place for predicting reactivity of proteins in aqueous solution at neutral pH?

2. Are these predominant sites of reactivity in a protein predictable, based on the primary amino acid sequences, and the regional hydropathy and flexibility near the reaction site? What percentage of reactive sites are not predictable based solely on sequence or hydrophobicity calculations?

3. Does the absolute local protein conformation play an overriding role in determining the reactivity of individual Asn, Gln, Asp and Met such that prediction of reaction "hot spots" based on primary sequence and hydrophobicity is a shot in the dark? Or is it just a subtle variable in the background, and other factors are predominant most of the time? There are examples in the literature where the local conformation and flexibility bring potential catalytic residues from distant regions in the sequence into close proximity of the deamidating amide side chain (Wright, 1991a). Alternatively, constraints on the backbone conformation may inhibit the deamidation of particular Asn residues (Kossiakoff, 1988). Further, potential catalytic side groups may be prevented from participating in the deamidation reaction because of hydrogen bonding or interactions with cofactors or ligands (Wright, 1991b). How much do these effects complicate the prediction of protein chemical reactivity?

4. Is this rate of chemical reaction fast enough to compromise a 2-year shelf life at 2–8°C and at pH 4.5–7.5? Although the kinetics of protein degradation are not addressed specifically in this report, it should be realized that all amino acids will degrade if followed long enough at sufficiently high temperatures, and the reader should be aware of this when reading the protein degradation literature (there are numerous examples of protein degradation at elevated temperatures and high or low pH, and these may not be representative of protein degradation in typical protein formulations).

2. PREDICTION OF PROTEIN CHEMICAL REACTIVITY BASED ON AMINO ACID SEQUENCE ANALYSIS

Although it has been known for years that certain amino acid sequences are prone to hydrolytic degradation (such as deamidation, cyclic imide formation, and iso-Asp formation at Asn, deamidation at Gln, or cyclic imide and iso-Asp formation at Asp), it has been argued that the neighboring substituent effects and conformational aspects are too complicated to allow routine prediction of chemical reactivity based on amino acid sequence and hydrophobicity/flexibility calculations. The same is believed to be true for Met oxidation; there is little correlation of reactivity and neighboring substituent effect (also called the sequence effect). To date, however, there does not exist a systematic analysis of protein reactivity in solution, such that a comparison of these studies is easily made. This chapter attempts to fill this need in formulation science, with the goal of attaining a better understanding of protein chemical reactivity in aqueous solution.

It is appropriate at this point to introduce the caveats in this analysis, lest the unwary reader be led astray from the main focus of this paper:

1. Proteins degrade by different pathways, both chemical and physical. The data and calculations herein do not address all protein degradation pathways, but only the chemical degradation pathways of deamidation, hydrolysis (cleavage), and oxidation. Degradation by other pathways including aggregation, precipitation, conformational denaturation, transamination, disulfide scrambling, reduction, enzymatic degradation, racemization, and other common routes are not part of this analysis. Further, there is no correction made for potential glycosylation at Asn (possible in the hot-spot motifs, -XNGS-, -XNGT-, -XNSS-, or -XNST-) which eliminates reaction at these potential hot spots.

2. There are several protein purification reports in which the isolated and purified protein is heterogeneous at a particular site, often Asn. The heterogeneity is usually caused by deamidation, giving Asp and iso-Asp. Many of these papers describe deamidation under extreme conditions that are not applicable to the long-term storage of protein formulations, including heating to 100°C, or acetic acid exposure during isolation. Further, these proteins are isolated from a biological milieu containing enzymes that may cause deamidation. There is sufficient evidence in the literature to suggest that deamidation can be significantly faster in a cellular or plasma medium than in aqueous solution of comparable pH and temperature (Nyberg *et al.*, 1985; O'Kelley *et al.*, 1985). Further, it is possible that the enzymatic deamidation pathway is different than the nonenzymatic pathway, so data generated under "work-up" conditions must be viewed cautiously.

3. Some proteins are quite small, such as secretin (27 amino acids) or insulin, and are close to the limit of being described as "large peptides." A few of these have been included in this analysis to thoroughly represent pharmaceutically relevant peptides and proteins, as well as to show that the data presented herein are directly applicable to smaller polypeptides as well.

4. Much of the literature on protein degradation focuses on determining the detailed mechanism of degradation and the factors that affect the reaction pathway(s). For example, the mechanistic distinctions in deamidation pathways have been studied in detail, in which deamidation occurs by cyclic imide formation, giving Asp or iso-Asp, or by deamidation of Asn, directly giving Asp without cyclic imide formation. This chapter does not attempt to review the excellent work in this area, but rather attempts to capitalize on it with the goal of addressing the sites of probable reaction and their likelihood of compromising the stability of a liquid protein formulation stored at 2–8°C for 1.5 years or more.

5. The "quality" of the different reports of protein degradation vary widely. Some studies are fairly extensive, for example, when conducted as part of a pharmaceutical drug development program. Others are short reports in the biochemical literature more than 20 years ago when the techniques for detecting protein degradation were not nearly as sophisticated as they are today. For example, detecting iso-Asp formation from Asp has been problematic by most chromatographic methods, and may be underreported in the protein degradation literature. The detection of other species, such as succinimide formation or a particular oxidized isoform, is also often difficult to detect and so may be underreported in older literature reports.

2.1. Common Chemical Degradation Pathways in Proteins

Much of our understanding of protein deamidation comes from the study of deamidation in small peptides. Several reviews on deamidation have been published (Robinson and Rudd, 1974; Wright, 1991b; Cleland *et al.*, 1993) and should be consulted if more detail is required. In general, deamidation is catalyzed by base,

heat, and ionic strength and is retarded by the addition of organic solvents (Capasso *et al.*, 1991). The rate of deamidation (as well as the detailed mechanism) is dictated by the pH and the adjacent amino acid(s). The deamidation rate for Asn is usually greater than for Gln, and is greatest when Asn or Gln are adjacent to Gly (-Asn-Gly- or -Gln-Gly-) (Robinson *et al.*, 1973a). The higher reactivity of the -Asn-Gly- bond compared to -Asn-X- (where X ≠ Gly) is shown by the degradation of Val-Tyr-Pro-Asn-X-Ala at pH 7.4 and 37°C. The half-lives for these peptides are X = Gly, 1.1; Ser, 8; Ala, 20; Leu, 70; Pro, 106 days, respectively. Hydrophobic or bulky amino acids in the sequence -Asn-X- appear to slow the deamidation rate considerably. At the preceding position, there are conflicting reports as to the nature of the substituent effect. Inspection of the rate data for peptides containing the -XNA- or the -XQA- motif shows that polar amino acids in the position -X-Asn- or -X-Gln- accelerate the deamidation rate, and bulky or hydrophobic residues tend to retard the deamidation rate. Figure 1 shows the substituent effect for the -X-Asn- and -X-Gln- motifs. In contrast, deamidation of peptides at pH 7.3 and 60°C containing the -XNS- motif showed no substituent effect (Tyler-Cross and Schirch, 1991). In this study, it is possible that any subtle substituent effect may be masked at the higher temperature of this reaction.

The model peptide containing the -GNA- motif was found to degrade exclusively via the cyclic imide intermediate from pH 5–12, and via direct hydrolysis of the amide side chain at acidic pH to give the Asp-hexapeptide (Patel and Borchardt, 1990). Under similar conditions, the deamidation half-lives for a series of pentapeptides yield values ranging from 6 days (Gly-Ser-Asn-His-Gly) to 3400 days (Gly-Thr-Gln-Ala-Gly) (Robinson *et al.*, 1973a; McKerrow and Robinson, 1974). At pH

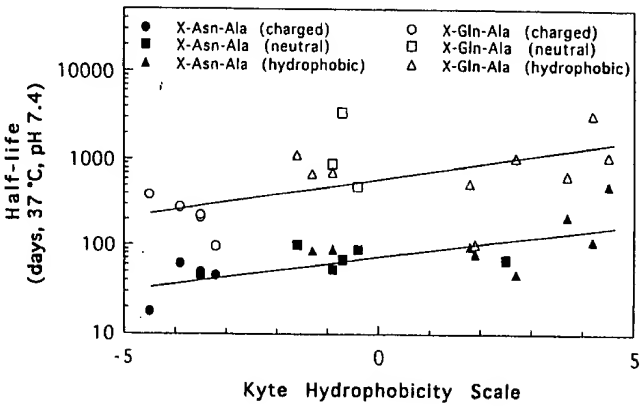
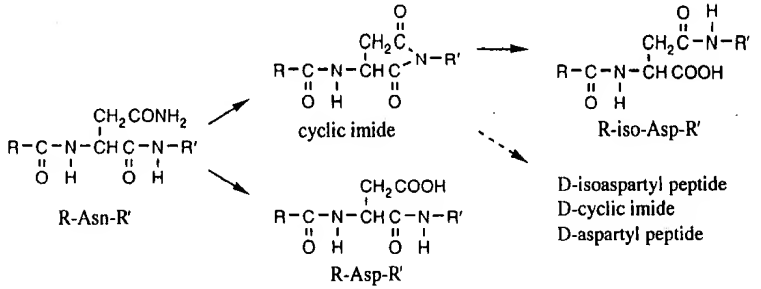


Figure 1. Correlation of deamidation half-life at pH 7.4 and 37°C with the Kyte–Doolittle hydrophobicity parameter. These data are from Robinson and Rudd (1974) and are determined by using a series of peptides defined by Gly-X-Asn-Ala-Gly. Inspection of the data show that, for both Asn and Gln, polar and charged amino acids adjacent to the reaction site accelerate the reaction rate, whereas hydrophobic or bulky residues decrease the rate of deamidation.

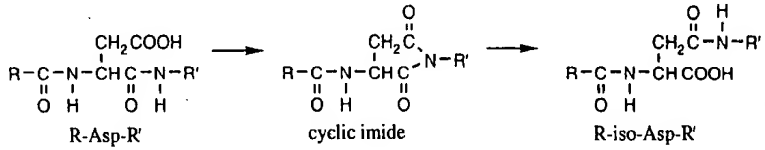
7.4 and 37°C, the rate of -Asn-Gly- bond cleavage was found to be 30- to 40-fold faster than for -Asp-Gly- (see below). A summary mechanism for Asn deamidation is shown in Scheme 2, including direct hydrolysis of the amide side chain and cyclic



Scheme 2

imide formation. This reaction may also result in racemization, thus forming the D-amino acid analogues.

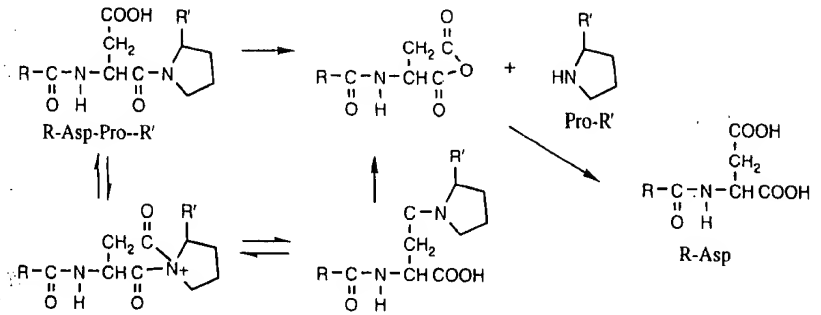
The -Asp-Gly- bond is also fairly reactive at neutral pH, yielding reversible isomerization between the Asp and iso-Asp forms via the cyclic imide intermediate (Scheme 3). Several Asp-containing peptides also yield detectable amounts of this



Scheme 3

intermediate (Bodansky *et al.*, 1967). The higher reactivity of the -Asp-Gly- bond is observed in the degradation of Val-Tyr-Pro-Asp-X-Ala at pH 7.4 and 37°C. The half-lives for these peptides are X = Gly, 41; Ser, 168; Ala, 266 days (Stephenson and Clarke, 1989). Iso-Asp also forms from Asp when Asp is adjacent to sterically hindered groups, such as in glucagon (-Asp-Tyr-) (Ota *et al.*, 1987) and calmodulin (-Asp-Gln-, -Asp-Thr-) (Ota and Clarke, 1989). Oliyai *et al.* (Oliyai and Borchardt, 1993) determined the effect of pH on the degradation of a model hexapeptide, in which the rate constant for -Asp-Gly- hydrolysis below pH 3 at 37°C was 7.5×10^{-4} $M^{-1}s^{-1}$, corresponding to a shelf life at pH 5 of approximately 0.5 year.

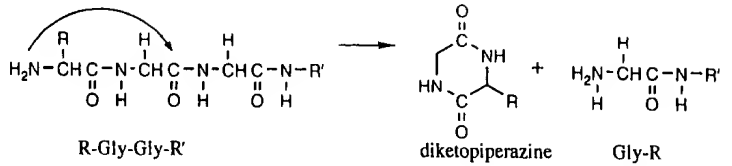
Asp is also reactive under acid conditions if the adjacent amino acid is proline, as in -Asp-Pro- (Schultz, 1967). For example, the reaction half-life of the -Asp-X- peptide bond in 0.015 N HCl at 110°C is much more rapid for Pro than for other amino acids: X = Pro, 11; Leu, 84; Ser, 108; Phe, 130; Lys, 228 min. The enhanced rate of this hydrolytic reaction is due to the increased leaving-group ability of the protonated proline due to the higher basicity of the proline nitrogen (Scheme 4). Model peptide studies suggest that this reaction is not sufficiently rapid at pH 5–7 and 2–8°C to compromise an aqueous-based protein formulation, but one should pay attention to



Scheme 4

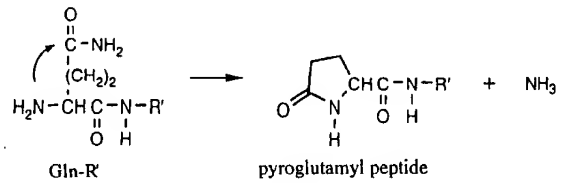
this degradation reaction as it is unlikely that the clipped fragments of the parent are biologically active.

There are other hydrolytic reactions that may compromise protein shelf life at pH 5–7, such as diketopiperazine (DKP) and pyroglutamic acid formation (Steinberg and Bada, 1983). Peptides containing glycine as the third amino acid from the N-termini undergo DKP formation much more easily than peptides with other amino acids in the third position (Sepetov *et al.*, 1991). Further, DKP formation is enhanced by incorporation of Pro or Gly into positions 1 or 2, whereas cyclization is completely prevented by blocking the α-amino group. Unfortunately, there is a paucity of data for this reaction, especially at 2–8°C, making a stability prediction difficult. It has been shown that there is a modest substituent effect at position 1 for DKP formation; reaction of X-Pro-Ala-Arg-Ser-Pro-Ser-Thr at 55°C and pH 7.0 for 3 days showed variable amounts of N-terminal degradation for X = Ala (83%), X = Val (35%), and X = Ser (89%) (Patel and Gitlin, 1995). In the same study it was shown that the pH of maximum stability for DKP formation of the Ala-Pro-Ala- peptide was approximately pH 4.5. The mechanism of DKP formation involves the nucleophilic attack of the N-terminal nitrogen on the amide carbonyl between the second and third amino acids (Scheme 5).



Scheme 5

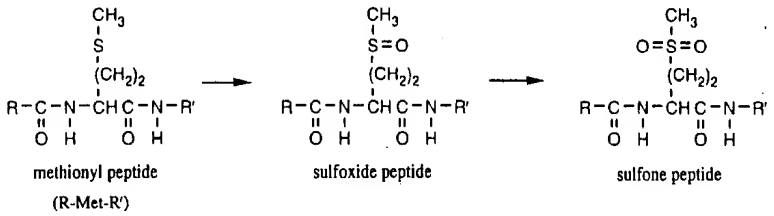
The reaction of N-terminal Gln is faster than predicted based on other amino acids, including Asn (Blomback, 1967). In this case, the Gln-amide undergoes nucleophilic attack by the N-terminal amino group, giving pyroglutamic acid (Scheme 6). Fukawa has shown that Gln-Gly reacts much faster than the other similar peptides studied, including Pro-Gln-Gly and Leu-Gln-Gly (Fukawa, 1967). Again, there are several kinetic studies of pyroglutamic acid formation at higher tempera-



Scheme 6

tures, but few at 2–8°C. The available data in small peptides however, may model the reaction rates of pyroglutamic acid seen in proteins with the Gln-X-Gly- N-terminal sequence, in that proteins often show flexible and disordered N-terminal sequences with little secondary structure. Another interesting reaction of glutamate has been observed for the chimeric Fab antibody fragment, ReoPro, wherein incubation of this protein at 37°C and pH 7.2 gives pyroglutamate, as identified by IEF and hydrophobic interaction chromatography (Everett *et al.*, 1995). The formation of pyroglutamic acid should not be universally considered a "degradation product," as nature has protected several proteins from aminopeptidase attack by this modification.

Another major degradation route for proteins in liquid formulations is thermal oxidation. The terminology "thermal" protein oxidation is actually a misnomer, as the degradation rate is often governed by trace amounts of peroxide, metal ions, light, base, and free-radical initiators (Johnson and Gu, 1988). Although there are several reactive amino acids that are known to oxidize (Met, Cys, cystine, His, Trp, and Tyr), a review of the literature shows that, under mild oxidative conditions at pH 5–7, Met is the predominant amino acid that oxidizes (Stadtman, 1990). Met oxidizes by both chemical and photochemical pathways to give methionine sulfoxide and, under extremely oxidative conditions (rarely found in protein pharmaceutical formulations), methionine sulfone (Scheme 7).



Scheme 7

Even though a great deal is known about reactive oxygen species, the presence (or absence) of these initiators makes the prediction of autooxidation in parenteral formulations imprecise. For example, free-radical oxidation involves the separate effects of initiation, propagation, and termination. Further, there are several reactive oxygen species including singlet oxygen $^1O_2$, superoxide radical $O_2^{\cdot-}$, alkyl or hydrogen peroxide ROOH or H_2O_2, hydroxyl radicals (HO· or HOO·), and halide oxygen complexes (ClO^-) (Halliwell and Gutteridge, 1990). There is limited published data on the oxidation of proteins in pharmaceutical formulations because only a few of the

proteins developed thus far have shown significant amounts of oxidation. Methionine residues in polypeptides show widely varying reactivity, as some Met residues are protected from oxidation by steric effects or inaccessibility, being buried in the hydrophobic core of the protein (Teh *et al.*, 1987). The second-order rate constants ($M^{-1}s^{-1}$) of Met oxidation by hydrogen peroxide have been determined at room temperature for Met free amino acid (0.93), Ac-Ser-Trp-Met-Glu-Glu-$CONH_2$ (1.07), Ac-$CysNH_2$-S-S-AcCys-Gly-Met-Ser-Thr-$CONH_2$ (1.0), and the Met in relaxin B chain at positions 25 and 4 (Met B^{25}, 0.85; Met B^4, 0.34) (Nguyen *et al.*, 1993a). This study shows that the peroxide-catalyzed degradation of Met has little temperature dependence (ΔH ~ 10–12 kcal/mol) and is negligibly effected by pH or ionic strength. The amount of peroxide in some excipients such as polyethylene glycols and surfactants varies widely (Hamburger *et al.*, 1975; McGinity *et al.*, 1975) and should be used cautiously in the formulation of Met-containing proteins. Using the data of Nguyen *et al.* (1993a), it is estimated that 1 nM peroxide in a Met-containing formulation would shorten the shelf life to less than 2 years.

2.2. Calculation of Protein Hydropathy and Flexibility

The general literature, and Genentech's GenBank data base, were surveyed for proteins that exhibit deamidation, hydrolysis, cyclization, or oxidation. Also included are a few unpublished observations from reliable laboratories. The protein sequences were scanned for the reactive residues, Asn, Asp, Gln, and Met, and the motifs surrounding these residues were tabulated as "reactive sites," although it is recognized that not all Asn, Asp, Gln, or Met are predicted to be reactive. To aid the reader, only the highly reactive motifs were labeled on the hydroflex plots (see below), and these included Asn-Gly, Asn-Ser, Asp-Gly, Gln-Gly, Asp-Pro, and Met.

The primary amino acid sequences were then used to construct "hydroflex" plots, consisting of the calculated hydropathy of the amino acid sequence, as well as its flexibility (see below). Hydropathy has been used to calculate antigenic determinants, as well as the surface characteristic of proteins (Hopp and Woods, 1981, 1983; Hopp, 1985, 1986). The hydropathy plot was constructed using the "hydro" program that scans the protein (or actually the individual hydropathy values assigned to each amino acid in the protein) with a window of specified size and computes the average hydrophobicity of each window (Watanabe, 1991). For example, a model protein shown in Scheme 8 is subjected to a window size of six amino acids, and the average hydropathy (ϕ) calculated.

Using this nomenclature, a hydropathy plot is simply a plot of ϕ versus amino acid number for the entire amino acid sequence. A window of 6 was chosen for several reasons: a window size of approximately 7–10 is believed to be optimal for searching for interior hydrophobic and exterior hydrophilic regions. A window size of 6–7 is believed to be optimal for searching for antigenic regions. Windows of sizes 5, 6, 7, and 10 amino acids were tested for several proteins with little visual difference

```
_____ϕ1        (ϕ1 = hydropathy average of ASDFGH, plotted at position 3)
_____ϕ2        (ϕ2 = hydropathy average of SDFGHC, plotted at position 4)
_____ϕ3        (ϕ3 = hydropathy average of DFGHCM, plotted at position 5)
_____ϕ4        (ϕ4 = hydropathy average of FGHCMN, plotted at position 6)
_____ϕ5        (ϕ5 = hydropathy average of GHCMNQ, plotted at position 7)
ASDFGHCMNQW...
123456789...
```

Scheme 8

in the hydropathy plots (data not shown). The optimal window size for the flexibility plot calculations is approximately 5 (see below), and so a window size of 6 amino acids was selected as a compromise between the methods for consistency. We chose the Kyte–Doolittle scale for our analysis of the various proteins. The individual amino acid hydropathy values used (Kyte–Doolittle parameters) are shown in Table I along with several other hydropathy scales for comparison.

For our purposes, the absolute values of the hydropathy values ϕ shown in the plots do not have significance; only the relative scale is important. Large positive values of ϕ denote regions of predicted high hydrophobicity; large negative values of ϕ denote regions of hydrophilicity. Although it is likely that hydrophobic regions tend to be found near the core of the protein, this is only a generalization and cannot be held as absolute from a simple calculation (the X-ray crystal, or NMR solution structure are the definitive indicators which amino acids are found in the core and which are found on the exterior of the protein).

The flexibility plots were calculated in a similar fashion using the parameters of hydrophobicity and side-chain volume according to Ragone *et al.* (1989). In this case, the relative flexibility scales gave values ranging from 1000 to 3000 and required normalization to plot with the hydrophobicity values. The mean value of zero for the flexibility plots was determined by computing the average flexibility of all of the proteins in the GenBank database, and included a statistical correction for the relative amino acids available in nature. The individual amino acid flexibility values are shown in Table I. Again, the absolute flexibility values do not have significance, but only the relative position on the plot. In the normalized plots, regions of flexibility have values less than zero; constrained regions have large positive values.

Hydropathy plots were calculated for several proteins using the Kyte, Hopp, Engleman, and Eisenberg hydropathy scales; all gave similar plots regardless of the hydropathy scale used (data not shown). The flexibility plots were often quite different than the hydropathy plots, largely because they are the cross product of hydropathy and amino acid side-chain volume (a correlate of "flexibility). So as to contrast the two major ways to analyze primary sequence analysis, the Kyte hydropathy plot and the flexibility plot are shown together to compare and contrast these methods. Conveniently, these plots provide at a single glance a visual picture of the protein. Reactive regions are typically found in large negative values, and stable regions found in large positive values of both hydropathy and flexibility (using either scale).

Table I. Summary of Hydropathy and Flexilibity Values for Individual Amino Acids[a]

AA	Kyte	Hopp	Engelman	Eisenberg	Ragoné[b]
A (Ala)	1.8	0.5	1.6	0.62	−0.91
C (Cys)	2.5	1.0	2.0	0.29	−0.17
D (Asp)	−3.5	−3.0	−9.2	−9.0	−0.68
E (Glu)	−3.5	−3.0	−8.2	−0.74	−0.68
F (Phe)	2.7	2.5	3.7	1.19	1.37
G (Gly)	−0.4	0.0	1.0	0.48	−1.40
H (His)	−3.2	0.5	−3.0	−0.4	0.25
I (Ile)	4.5	1.8	3.1	1.38	1.09
K (Lys)	−3.9	−3.0	−8.8	1.50	0.13
L (Leu)	3.7	1.8	2.8	1.06	0.89
M (Met)	1.9	1.3	3.4	0.64	0.83
N (Asn)	−3.5	−0.2	−4.8	−0.78	−0.42
P (Pro)	−1.6	0.0	−0.2	0.12	−0.52
Q (Gln)	−3.5	−0.2	−4.1	−0.85	0.06
R (Arg)	−4.5	−3.0	12.3	−2.53	0.71
S (Ser)	−0.9	−0.3	0.6	−0.18	−1.01
T (Thr)	−0.7	0.4	1.2	−0.05	−0.58
V (Val)	4.2	1.5	2.6	1.08	0.52
W (Trp)	−0.9	3.4	1.9	0.81	2.00
Y (Tyr)	−1.3	2.3	−0.7	0.26	1.21

[a]Kyte and Doolittle (1982); Hopp and Woods (1981, 1983); Engelman *et al.* (1986); Eisenberg (1984); Ragoné *et al.* (1989).

[b]Corrected to provide a mean at ≃ 0 and maximum and minimum values over an average window of six amino acids of approximately +1.0 and −1.0, respectively.

3. SUMMARY OF PROTEIN STABILITY IN AQUEOUS SOLUTION

The strategy used herein is straightforward: (i) Assemble all of the data on degradation of proteins in pharmaceutical liquid formulations (or in model formulations consisting of aqueous solution of pH ~ 4.5–7.5) where there is evidence for degradation by hydrolysis, cyclization, deamidation, or oxidation. Include salient data obtained from protein purification studies (if controls are available showing reaction in aqueous solution). These data may have some peculiarities due to enzyme catalyzed reactions. (ii) Compile the relevant primary sequence information for these proteins, including a subset analysis of the reactive groups Asn, Asp, Gln, and Met. (iii) Analyze the primary sequence in terms of hydrophobicity and flexibility in order to "guesstimate" regions of preferred chemical reactivity. In this analysis it is assumed that these reactive regions are also hydrophilic, and thus have a higher probability of being on the "outside" of the protein, and should be fairly flexible so as to allow the correct geometry for reaction. The following pages are summaries of these parameters for different proteins. Included in each summary is the primary

sequence of the protein, the motifs for all Asn, Asp, Gln, and Met, the calculated hydroflex plot, a short summary of the degradation pathway(s) reported in the literature, and comments on how predictive the primary sequence and the hydroflex plot were for protein degradation in aqueous solution (albeit retrospectively). Further mention is noted as to the reliability of the data for prediction of peptide/protein degradation under neutral pH formulation conditions.

- **Adrenocorticotropin (ACTH) (39 residues)**

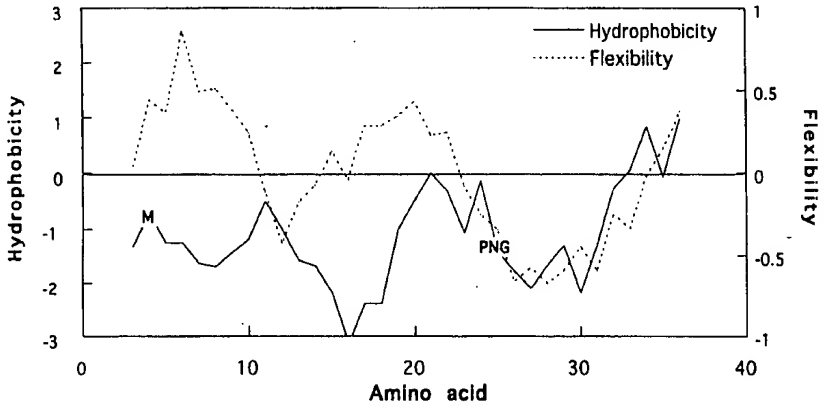
SEQUENCE

SYSMEHFRWGKPVGKKRRPVKVYPNGAEDESAEAFPLEF

REACTIVE SITES

.N.(1)	.D.(1)	.M.(1)	.Q.(0)
25 PNG	29 EDE	4 SME	

HYDROFLEX PLOT



PREDICTED REACTIVITY AND DEGRADATION OF ACTH

ACTH contains only a single site susceptible to hydrolytic degradation, Asn-25, with the -PNG- motif. Further, Asn-25 is located in a region predicted to be fairly flexible and hydrophilic, suggesting that this is the predominant reactive residue. Based on this, ACTH is expected to degrade primarily at Asn-25. Under neutral pH conditions, the 39-mer ACTH

peptide underwent deamidation at Asn-25 to give the cyclic imide, and the Asp-25 and iso-Asp-25 ACTH variants (Aswad, 1984; Patel, 1993). The reactivity of Asn-25 was studied extensively in the parent hormone and in smaller peptides of similar motif about Asn (such as the hexapeptide VTPNGA). No oxidation at Met was reported.

Agglutinin (171 residues)

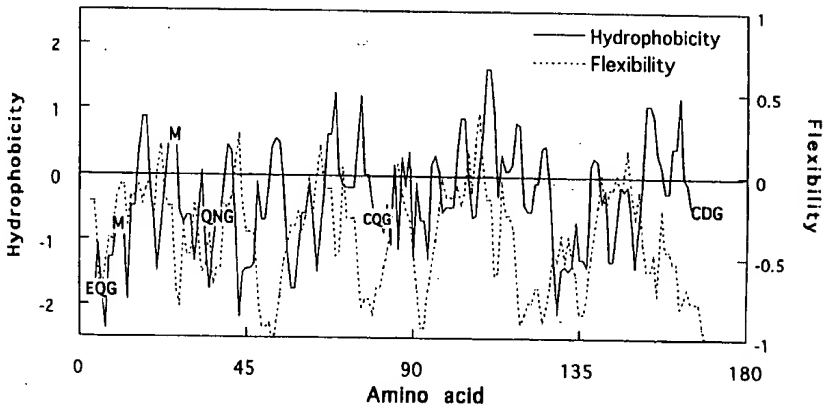
SEQUENCE

QRCGEQGSNMECPNNLCCSQYGYCGMGGDYCGKGCQNGACWTSKRCGSQAGGA-
TCTNNQCCSQYGYCGFGAEYCGAGCQGGPCRADIKCGSQAGGKLCPNNLCCSQW-
GFCGLGSEFCGGGCQSGACSTDKPCGKDAGGRVCTNNYCCSKWGSCGIGPGYCGA-
GCQSGGCDG

REACTIVE SITES

.N.(10)		.D.(5)	.M.(2)	.Q.(11)	
9 SNM	100 PNN	29 GDY	10 NME	6 EQG	79 CQG
14 PNN	101 NNL	86 ADI	26 GMG	20 SQY	92 SQA
15 NNL	143 TNN	129 TDK		36 CQN	106 SQW
37 QNG	144 NNY	135 KDA		49 SQA	122 CQS
57 TNN		170 CDG		59 NQC	165 CQS
58 NNQ				63 SQY	

HYDROFLEX PLOT



PREDICTED REACTIVITY AND DEGRADATION OF AGGLUTININ

The hydroflex plot shows that there are a few predicted sites of chemical reactivity, notably the Asn-Gly at position 37, and possibly the Asp-Gly at the C-termini. The nucleotide sequence code for agglutinin encodes for Asn at position 37 (within the -QNG- motif), yet amino acid sequence analysis indicated that Asp was the predominant amino acid observed at this site (Wright and Raikhel, 1989). It was not determined if this discrepancy was due to deamidation in the intact protein or in the proteolytically generated peptides used in the sequencing. No oxidation of Met was reported.

- **Aldolase (363 residues)**

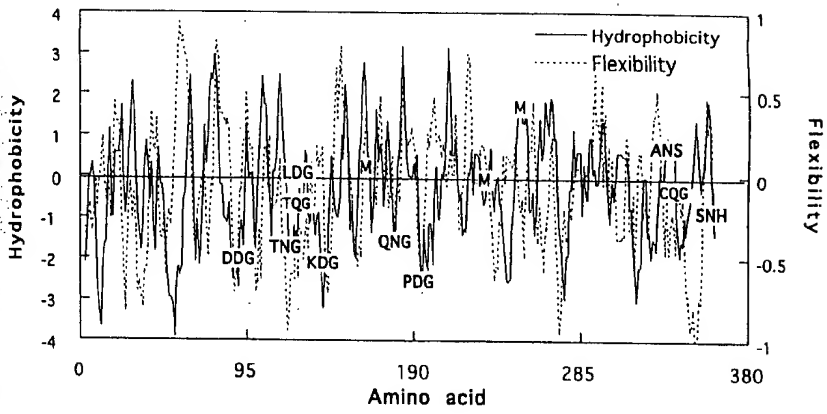
SEQUENCE

PHSHPALTPEQKKELSDIAHRIVAPGKGILAADESTGSIAKRLQSIGTENTEENRRFYR-
QLLLTADDRVNPCIGGVILFHETLYQKADDGRPFPQVIKSKGGVVGIKVDKGVVPLA-
GTNGETTTQGLDGLSERCAQYKKDGADFAKWRCVLKIGEHTPSALAIMENANVLA-
RYASICQQNGIVPIVEPEILPDGDHDLKRCQYVTEKVLAAVYKALSDHHIYLEGTLL-
KPNMVTPGHACTQKYSHEEIAMATVTALRRTVPPAVTGVTFLSGGQSEEEASINLNA-
INKCPLLKPWALTFSYGRALQASALKAWGGKKENLKAAQEEYVKRALANSLACQG-
KYTPSGQAGAAASESLFISNHAY

REACTIVE SITES

.N.(14)	.D.(14)	.M.(3)	.Q.(16)
50 ENT	17 SDI	164 IME	11 EQK
54 ENR	33 ADE	232 NMV	44 LQS
70 VNP	66 ADD	250 AMA	60 RQL
119 TNG	67 DDR		85 YQK
166 ENA	88 ADD		95 PQV
168 ANV	89 DDG		125 TQG
180 QNG	109 VDK		136 AQY
231 PNM	128 LDG		178 CQQ
282 INL	148 KDG		179 QQN
284 LNA	143 ADF		202 CQY
287 INK	193 PDG		241 TQK
319 ENL	195 GDH		274 GQS
334 ANS	197 HDL		306 LQA
360 SNH	218 SDH		324 AQE
			339 CQG
			347 GQA

HYDROFLEX PLOT



PREDICTED REACTIVITY AND DEGRADATION OF ALDOLASE

This molecule has several predicted degradation sites. Isolation of rabbit muscle aldolase and subsequent amino acid sequencing of the carboxyl-terminal peptide liberated by chymotrypsin hydrolysis shows that Asn undergoes deamidation to give Asp within the motif, -ISNHAY (Midelfort and Mehler, 1972). It has been pointed out that this Asn may be activated by the neighboring His, but otherwise the Asn-Ala motif is usually considered poorly reactive, as based on data obtained from small peptides. In fact, several other proteins have the -XNH- motif, including ARSP, anti-HER-2, 4D5 antibody, 17-1A antibody, CD4-IgG, chloroperoxidase, acidic-FGF, HXGT, IFN-β, OKT-3 antibody, SHMT, and t-PA, and showed no sign of reacting at this site, indicating that -XNH- is not particularly activating unless composed of -SNH-. In this study, no control experiments were carried out to show that the same deamidation reaction occurs in pH 4.5–7.5 buffer, and reaction at this site may be enzymatic in nature. Further, insufficient controls were carried out to determine if deamidation, cyclization, or oxidation occurred at many of the other sites predicted to be labile.

- **Amylin Antagonist (24 residues)**

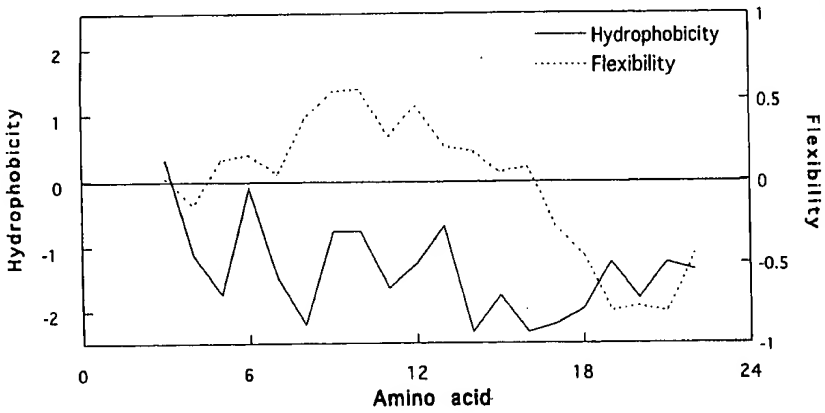
SEQUENCE

Ac-LGRLSQELHRLQTYPRTNTGSNTY-CONH$_2$

REACTIVE SITES

.N.(2)	.D.(0)	.M.(0)	.Q.(2)
18 TNT			6 SQE
22 SNT			12 LQT

HYDROFLEX PLOT



PREDICTED REACTIVITY AND DEGRADATION OF AMYLIN ANTAGONIST

This peptide is acylated at the N-termini and has the carbamoyl moiety at the C-termini (no account for these modifications was made in the hydroflex plot). Inspection of the hydroflex plot shows that this peptide should be quite stable in that it is devoid of the "traditional" hot spots for chemical degradation. Asn-18 and Asn-22 are adjacent to Thr, which is reported to activate reactivity at Asn only slightly. The solution stability of this amylin antagonist was investigated under acidic conditions (pH 2.6–5.0), approaching the desired pH range for parenteral formulations. Deamidation at Asn-22 was observed, with a rate minimum at pH 4.3, resulting in the formation of iso-Asp-22 and Asp-22 (3.2:2), consistent with cyclic imide formation. Deamidation at Asn-18 was not detected (Darrington, 1995).

- **Amyloid-Related Serum Protein (ARSP) (104 residues)**

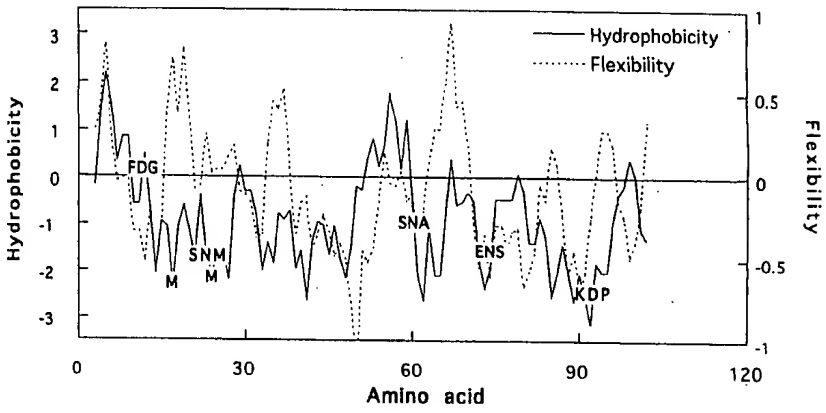
SEQUENCE

RSFFSFLGEAFDGARDMWRAYSNMREANYIGSDKYFHARGNYDAAKRGPGGAWA-AEVISNARENIQRFFGHDAENSLADQAANEWGRSGKDPNHFRPAGLPEKY

REACTIVE SITES

.N.(8)		.D.(7)		.M.(2)	.Q.(2)
23 SNM	64 ENI	12 FDG	72 HDA	17 DMW	66 IQR
28 ANY	75 ENS	16 RDM	79 ADQ	24 NMR	80 DQA
41 GNY	83 ANE	33 SDK	91 KDP		
60 SNA	93 PNH	43 YDA			

HYDROFLEX PLOT



PREDICTED REACTIVITY AND DEGRADATION OF AMYLOID-RELATED SERUM PROTEIN

Isolation of amyloid-related serum protein (ARSP) gives a 104-amino-acid protein that shows microheterogeneity at Asn-23 (-SNM-), Asn-60 (-SNA-), and Asn-75 (-ENS-), where only the later motif (-XNS-) is predicted to be chemically reactive at neutral pH (Sletten *et al.*, 1983). No controls were carried out to determine if this deamidation was due to isolation or differences in protein expression from different patients. Insufficient data was presented to allow the estimation of degradation at pH 4.5–7.5 at 5°C.

- **Angiogenin (123 residues)**

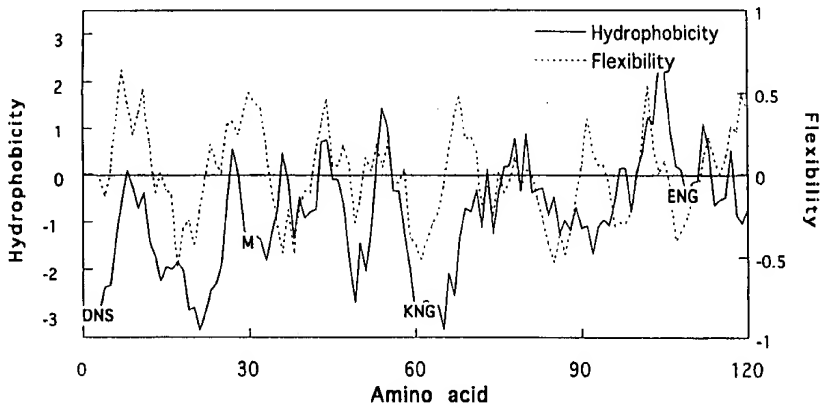
SEQUENCE

EDNSRYTHFLTQHYDAKPQGRDDRYCESIMRRRGLTSPCKDINTFIHGNKRSIKAICE-NKNGNPHRENLRISKSSFQVTTCKLHGGSPWPPCQYRATAGFRNVVVACENGLPVH-LDQSIFRRP

REACTIVE SITES

.N.(9)		.D.(6)	.M.(1)	.Q.(5)
3 DNS	68 ENL	2 END	30 IMR	12 TQH
43 INT	102 RNV	15 YDA		19 PQG
49 GNK	109 ENG	22 RDD		77 FQV
59 ENK		23 DDR		93 CQY
61 KNG		41 KDI		117 DQS
63 GNP		116 LDQ		

HYDROFLEX PLOT



PREDICTED REACTIVITY AND DEGRADATION OF ANGIOGENIN

The hydroflex plot shows that Asn-61 and Asn-109 are likely spots for reactivity, in that they are located adjacent to Gly and are found in moderately flexible regions. A third reactive site could also be Asn-3 in the -DNS- motif. Incubation of angiogenin at pH 8 and 4°C for 2 years resulted in approximately 35% loss of the original molecule (Hallahan *et al.*, 1992). Degradation occurred simultaneously at Asn-61 (-KNG-) and Asn-109 (-ENG-), which likely accounts for their observation of a third (and unidentified) acidic product—the doubly deamidated molecule. Alternatively, deamidation may have occurred at Asn-3, in that the reaction product of this third reaction product was not identified. Deamidation resulted in a dramatic loss in biological activity.

- **Anti-HER-2 Heavy Chain (450 residues)**

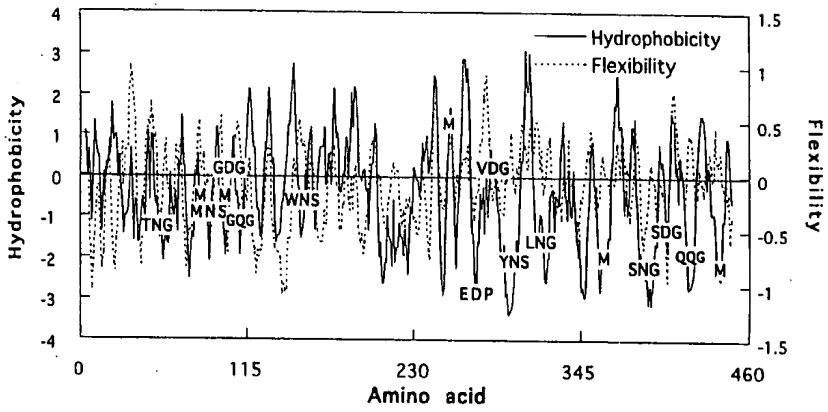
SEQUENCE

EVQLVESGGGLVQPGGSLRLSCAASGFNIKDTYIHWVRQAPGKGLEWVARIYPTNG-
YTRYADSVKGRFTISADTSKNTAYLQMNSLRAEDTAVYYCSRWGGDGFYAMDYW-
GQGTLVTVSSASTKGPSVFPLAPSSKSTSGGTAALGCLVKDYFPEPVTVSWNSGALT-
SGVHTFPAVLQSSGLYSLSSVVTVPSSSLGTQTYICNVNHKPSNTKVDKKVEPKSCD-
KTHTCPPCPAPELLGGPSVFLFPPKPKDTLMISRTPEVTCVVVDVSHEDPEVKFNWY-
VDGVEVHNAKTKPREEQYNSTYRVVSVLTVLHQDWLNGKEYKCKVSNKALPAPIE-
KTISKAKGQPREPQVYTLPPSREEMTKNQVSLTCLVKGFYPSDIAVEWESNGQPEN-
NYKTTPPVLDSDGSFFLYSKLTVDKSRWQQGNVFSCSVMHEALHNHYTQKSLSLS-
PGK

REACTIVE SITES

.N.(19)	.D.(18)	.M.(5)	.Q.(16)
28 FNI	31 KDT	83 QMN	3 VQL
55 TNG	62 ADS	107 AMD	13 VQP
77 KNT	73 ADT	255 LMI	39 RQA
84 MNS	90 EDT	361 EMT	82 LQM
162 WNS	102 GDG	431 VMH	112 GQG
204 CNV	108 MDY		178 LQS
206 VNH	151 KDY		199 TQT
211 SNT	215 VDK		298 EQY
279 FNW	224 CDK		314 HQD
289 HNA	252 KDT		345 GQP
300 YNS	268 VDV		350 PQV
318 LNG	273 EDP		365 NQV
328 SNK	283 VDG		389 GQP
364 KNQ	315 QDW		421 WQQ
387 SNG	379 SDI		422 QQG
392 ENN	402 LDS		441 TQK
393 NNY	404 SDG		
424 GNV	416 VDK		
437 HNH			

HYDROFLEX PLOT



PREDICTED REACTIVITY AND DEGRADATION OF ANTI-HER-2 ANTIBODY HEAVY CHAIN

Inspection of the amino acid sequence for the anti-HER-2 heavy chain shows that there are several reactive sites, including predicted deamidation at Asn-318 in the -LNG- motif, Asn-387 in the -SNG- motif, iso-Asp formation at Asn-55 in the -TNG- motif, at Asp-102 in the

-GDG- motif, at Asp-283 in the -VDG- motif, and Asp-404 in the -SDG- motif. This antibody is formulated as a liquid in 5 mM isotonic acetate, pH 5.0, 0.01% Polysorbate 20. After 1.5 years at 2–8°C storage, it was shown using an ion-exchange assay the formation of cyclic imide at Asp-102 (located in the CDR3 region); this identification of succinimide intermediate was done also carried out by HIC after alkaline hydroxylamine cleavage (Kwong and Harris, 1985). This degradation product, as well as the iso-Asp product (Harris *et al.*, 1995), has been isolated and shown to retain full biological activity. All other assays were virtually unchanged after storage at 2–8°C. Although this protein did not show oxidation under formulation conditions at 2–8°C, rapid oxidation of Met-255 and Met-431 was catalyzed by *t*-butylhydroperoxide (Shen *et al.*, 1996).

• Anti-HER-2 Light Chain (214 residues)

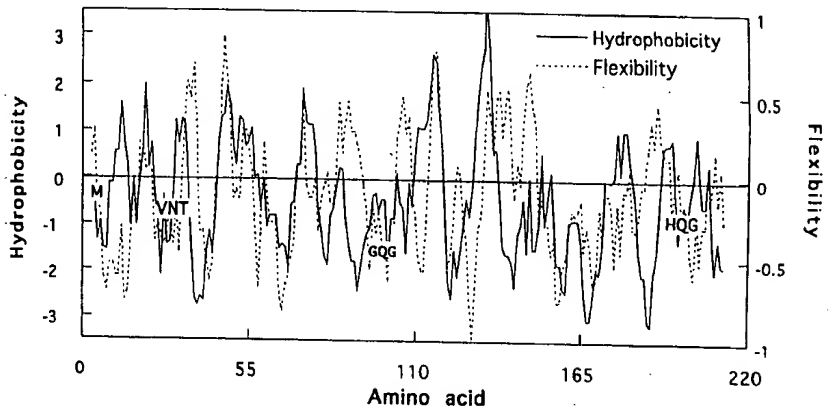
SEQUENCE

DIQMTQSPSSLSASVGDRVTITCRASQDVNTAVAWYQQKPGKAPKLLIYSASFLYSG-
VPSRFSGSRSGTDFTLTISSLQPEDFATYYCQQHYTTPPTFGQGTKVEIKRTVAAPSVF-
IFPPSDEQLKSGTASVVCLLNNFYPREAKVQWKVDNALQSGNSQESVTEQDSKDST-
YSLSSTLTLSKADYEKHKVYACEVTHQGLSSPVTKSFNRGEC

REACTIVE SITES

.N.(6)	.D.(9)	.M.(1)	.Q.(15)
30 VNT	17 GDR	4 QMT	3 IQM
137 LNN	28 QDV		6 TQS
138 NNF	70 TDF		27 SQD
152 DNA	82 EDF		37 YQQ
158 GNS	122 SDE		38 QQK
210 FNR	151 VDN		79 LQP
	167 QDS		89 CQQ
	170 KDS		90 QQH
	185 ADY		100 GQG
			124 EQL
			147 VQW
			155 LQS
			160 SQE
			166 EQD
			199 HQG

HYDROFLEX PLOT



PREDICTED REACTIVITY AND DEGRADATION OF ANTI-HER-2 ANTIBODY LIGHT CHAIN

Inspection of the amino acid sequence for the anti-HER-2 light chain shows that there are few reactive sites, perhaps the most reactive being the single Met. Gln-Gly appears in the HER-2 light chain, but is the least reactive of the traditional (Asn-Gly, Asn-Ser, Asp-Gly, Asp-Pro, Met, and Gln-Gly) hot spots. This absence of hot spots suggests that the light chain of anti-HER-2 should be fairly stable compared to the heavy chain. Some deamidation of the light chain has been observed at Asn-30 in CDR1 of the light chain during the cell culture process, typically 10–12%. Deamidation of Asn-30 in one chain resulted in an ~18% decrease in activity as measured in the ECD plate binding assay (Harris, 1995; Shire, 1995), but little has been observed at pH 5. This residue is not a traditional hot spot, but is predicted to be in a flexible hydrophilic region. All other assays are virtually unchanged at 2–8°C storage.

Antibody 4D5 Heavy Chain (450 residues)

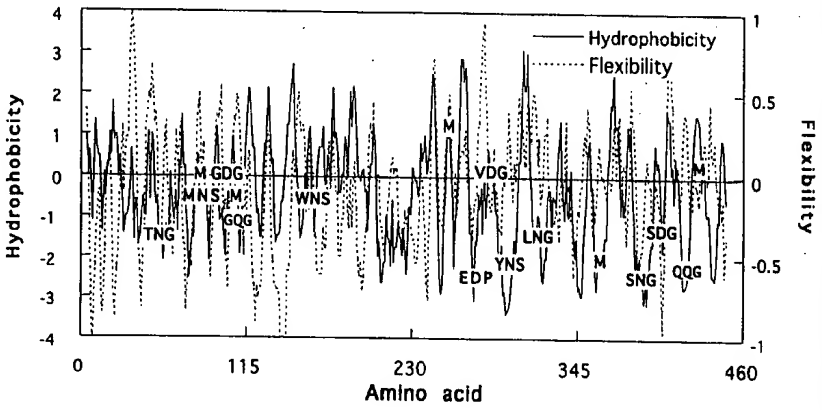
SEQUENCE

EVQLVESGGGLVQPGGSLRLSCAASGFNIKDTYIHWVRQAPGKGLEWVARIYPTNG-
YTRYADSVKRFTISADTSKNTAYLQMNSLRAEDTAVYYCSRWGGDGFYAMDYWG-
QGTLVTVSSASTKGPSVFPLAPSSKSTSGGTAALGCLVKDYFPEPVTVSWNSGALTS-
GVHTFPAVLQSSGLYSLSSVVTVPSSSLGTQTYICNVNHKPSNTKVDKKVEPKSCDK-
THTCPPCPAPELLGGPSVFLFPPKPKDTLMISRTPEVTCVVVDVSHEDPEVKFNWYV-
DGVEVHNAKTKPREEQYNSTYRVVSVLTVLHQDWLNGKEYKCKVSNKALPAPIEK-
TISKAKGQPREPQVYTLPPSREEMTKNQVSLTCLVKGFYPSDIAVEWESNGQPENNY-
KTTPPVLDSDGSFFLYSKLTVDKSRWQQGNVFSCSVMHEALHNHYTQKSLSLSPGK

REACTIVE SITES

.N.(19)	.D.(18)	.M.(5)	.Q.(16)
28 FNI	31 KDT	83 QMN	3 VQL
55 TNG	62 ADS	107 AMD	13 VQP
77 KNT	73 ADT	255 LMI	39 RQA
84 MNS	90 EDT	361 EMT	82 LQM
162 WNS	102 GDG	431 VMH	112 GQG
204 CNV	108 MDY		178 LQS
206 VNH	151 KDY		199 TQT
211 SNT	215 VDK		298 EQY
279 FNW	224 CDK		314 HQD
289 HNA	252 KDT		345 GQP
300 YNS	268 VDV		350 PQV
318 LNG	273 EDP		365 NQV
328 SNK	283 VDG		389 GQP
364 KNQ	315 QDW		421 WQQ
387 SNG	379 SDI		422 QQG
392 ENN	402 LDS		441 TQK
393 NNY	404 SDG		
424 GNV	416 VDK		
437 HNH			

HYDROFLEX PLOT



PREDICTED REACTIVITY AND DEGRADATION OF 4D5 ANTIBODY HEAVY CHAIN

There are a number of predicted potentially reactive deamidation and isomerization sites in the 4D5 heavy chain. Inspection of the primary amino acid sequence for the 4D5 heavy chain shows that the most reactive is predicted to be Asn-55 within the -TNG- motif in the

CDR2 domain, Asn-318 within the -LNG- motif, Asn-387 within the -SNG- motif. All reside in a region predicted to be hydrophilic and flexible. There are several other hot spots, including several Asp-Gly (that may isomerize to form iso-Asp-Gly), and three Met residues. The Asp-Gly within the -GNG- motif may be particularly labile, as aspartic acid residues adjacent to C-terminal glycine residues are very susceptible to imide formation and subsequent isomerization (Cleland *et al.*, 1993). This antibody was found to be stable for more than 12 months at 2–8°C by all methods tested, and showed some loss in activity at 25°C and 40°C (Harris, 1995; Shire, 1995). Decrease in activity did not correlate with formation of aggregates, but appears related to alterations in the protein which result in the generation of acidic bands as detected by IEF. No conclusive identification of the reactive site in the heavy chain (if at all) was made.

Antibody 4D5 Light Chain (214 residues)

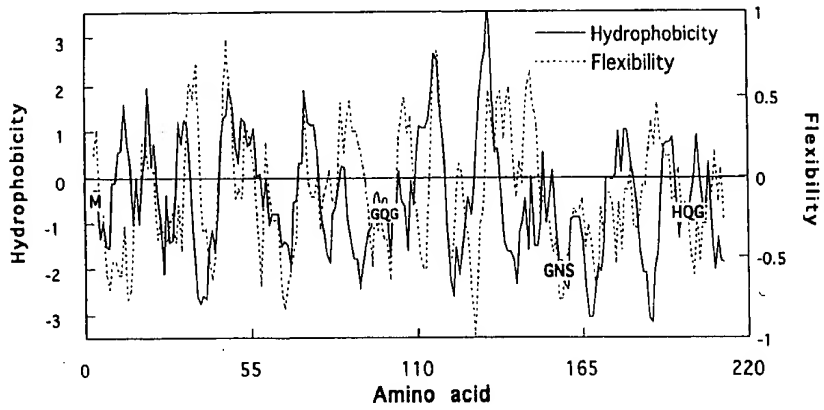
SEQUENCE

DIQMTQSPSSLSASVGDRVTITCRASQDVNTAVAWYQQKPGKAPKLLIYSASFLYSG-
VPSRFSGSRSGTDFTLTISSLQPEDFATYYCQQHYTTPPTFGQGTKVEIKRTVAAPSVF-
IFPPSDEQLKSGTASVVCLLNNFYPREAKVQWKVDNALQSGNSQESVTEQDSKDST-
YSLSSTLTLSKADYEKHKVYACEVTHQGLSSPVTKSFNRGEC

REACTIVE SITES

.N.(6)	.D.(9)	.M.(1)	.Q.(15)
30 VNT	17 GDR	4 QMT	3 IQM
137 LNN	28 QDV		6 TQS
138 NNF	70 TDF		27 SQD
152 DNA	82 EDF		37 YQQ
158 GNS	122 SDE		38 QQK
210 FNR	151 VDN		79 LQP
	167 QDS		89 CQQ
	170 KDS		90 QQH
	185 ADY		100 GQG
			124 EQL
			147 VQW
			155 LQS
			160 SQE
			166 EQD
			199 HQG

HYDROFLEX PLOT



PREDICTED REACTIVITY AND DEGRADATION OF 4D5 ANTIBODY LIGHT CHAIN

Inspection of the primary amino acid sequence for the 4D5 light chain shows that there are few reactive sites, perhaps the most reactive being Asn-158 within the Asn-Ser motif, as it resides in a predicted hydrophilic and flexible region. None of the hot spots reside in the CDR domain. This antibody, formulated as a liquid in isotonic 5 mM acetate at pH 5.0 with 0.01% polysorbate 20, was stable for more than 12 months at 2–8°C (Shire, 1995). At 25°C and 40°C there were decreases in activity (up to 77%). The decrease in activity did not correlate with formation of aggregates, but appeared to be related to alterations in the protein which result in the generation of acidic bands as detected by IEF. No conclusive identification of the reactive site in the light chain (if at all) was made.

- **Antibody 17-1A Heavy Chain (446 residues)**

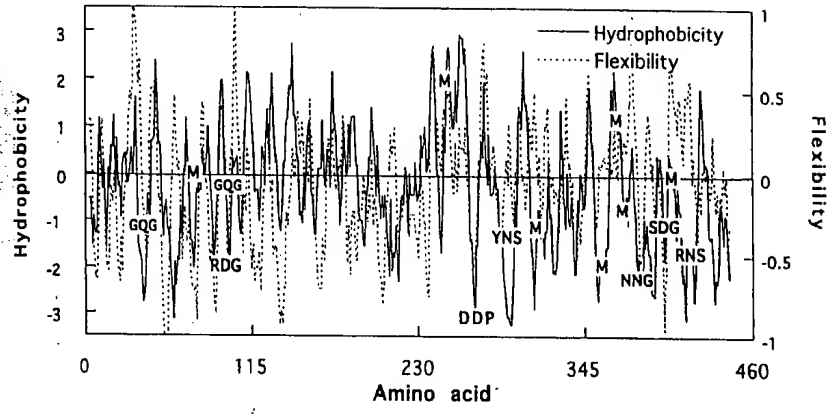
SEQUENCE

QVQLQQSGAELVRPGTSVKVSCKASGYAFTNYLIEWVKQRPGQGLEWIGVINPGSG-
GTNYNEKFKGKATLTADKSSSTAYMQLSSLTSDDSAVYFCARDGPWFAYWGQGTLV-
TVSAAKTTAPSVYPLAPVCGDTTGSSVTLGCLVKGYFPEPVTLTWNSGSLSSGVHTF-
PAVLQSDLYTLSSSVTVTSSTWPSQSITCNVAHPASSTKVDKKIEPRGPTIKPCPPCKC-
PAPNLLGGPSVFIFPPKIKDVLMISLSPIVTCVVDVSEDDPDVQISWFVNNVEVHTAQ-
TQTHREDYNSTLRVVSALPIQHQDWMSGKEFKCKVNNKDLPAPIERTISKPKGSVR-
APQVYVLPPPEEEMTKKQVTLTCMVTDFMPEDIYVEWTNNGKTELNYKNTEPVLD-
SDGSYFMYSKLRVEKKNWVERNSYSCSVVHEGLHNHHTTKSFSRTPGK

REACTIVE SITES

.N.(19)				.D.(19)				.M.(7)		.Q.(16)	
31	TNY	323	VNN	73	ADK	269	DDP	81	YMQ	3	VQL	287	AQT
52	INP	324	NNK	89	SDD	271	PDV	251	LMI	5	LQQ	289	TQT
59	TNY	382	TNN	90	DDS	294	EDY	313	WMS	6	QQS	308	IQH
61	YNE	383	NNG	99	RDG	311	QDW	357	EMT	39	KQR	310	HQD
159	WNS	389	LNY	133	GDT	326	KDL	367	CMV	43	GQG	346	PQV
199	CNV	392	KNT	176	SDL	370	TDF	372	FMP	82	MQL	361	KQV
232	PNL	415	KNW	210	VDK	375	EDI	405	FMY	108	GQG		
279	VNN	420	RNS	248	KDV	398	LDS			174	LQS		
280	NNV	433	HNH	264	VDV	400	SDG			194	SQS		
296	YNS			268	EDD					273	VQI		

HYDROFLEX PLOT



PREDICTED REACTIVITY AND DEGRADATION OF 17-1A ANTIBODY HEAVY CHAIN

The primary amino acid sequence for the 17-1A antibody heavy chain shows that Asn-383 should be very reactive in that it resides within the -NNG- motif, located in a hydrophilic and flexible region. There are several other reactive sites, including Asp-Pro, Asp-Gly, and Met. The observed reaction of the 17-1A antibody occurred largely at the C-terminus, with loss of Lys in a nonenzymatic process (enzyme inhibitors had no effect on this process, and there was no C-terminal reaction of other antibodies sensitive to C-terminal clipping when incubated with 17-1A antibody). This reaction pathway was found to be stabilized at acid pH. It is likely that this novel pathway was not the only reaction pathway for the 17-1A antibody, as numerous IEF bands were observed over time. This protein does, however, represent another example of an "unexpected" protein reaction at a non-hot-spot site.

Antibody 17-1A Light Chain (214 residues)

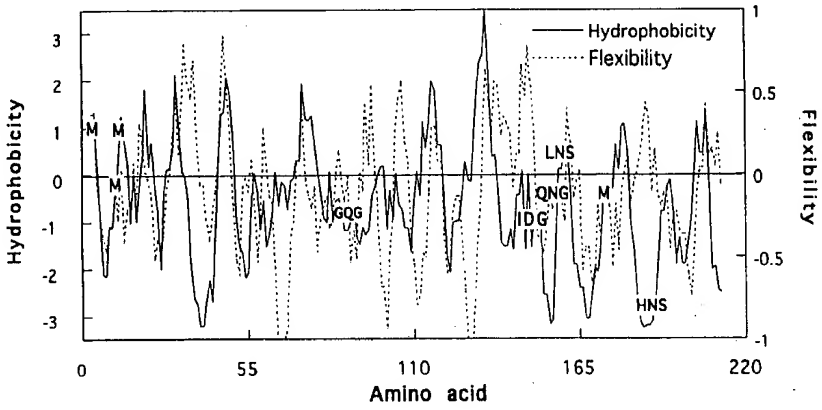
SEQUENCE

NIVMTQSPKSMSMSVGERVTLTCKASENVVTYVSWYQQKPEQSPKLLIYGASNRYT-
GVPDRFTGSGSATDFTLTISSVQAEDLADYHCGQGYSYPYTFGGGTKLEIKRADAAP-
TVSIFPPSSEQLTSGGASVVCFLNNFYPKDINVKWKIDGSERQNGVLNSWTDQDSK-
DSTYSMSSTLTLTKDEYERHNSYTCEATHKTSTSPIVKSFNRNEC

REACTIVE SITES

.N.(10)	.D.(11)	.M.(4)	.Q.(9)
28 ENV	60 PDR	4 VMT	6 TQS
53 SNR	70 TDF	11 SMS	37 YQQ
137 LNN	82 EDL	13 SMS	38 QQK
138 NNF	85 ADY	175 SMS	42 EQS
145 INV	110 ADA		79 VQA
157 QNG	143 KDI		90 GQG
161 LNS	151 IDG		124 EQL
190 HNS	165 TDQ		156 RQN
210 FNR	167 QDS		166 DQD
212 RNE	170 KDS		
	184 KDE		

HYDROFLEX PLOT



PREDICTED REACTIVITY AND DEGRADATION OF 17-1A ANTIBODY LIGHT CHAIN

The primary amino acid sequence for the 17-1A antibody light chain shows that Asn-157 should be reactive in that it resides within the -QNG- motif, although its motif is located in a region of only modest hydrophilicity and predicted chain flexibility. There are several other reactive sites, including Asp-Gly and Met. The degradation of the 17-1A antibody occurred largely on the heavy chain (see previous entry), although the authors observed that several new IEF bands were found over time, supportive of possible reaction at this Asn-Gly hot spot (Everett, 1995). No oxidation of Met was reported.

Antibody E25 Light Chain (218 residues)

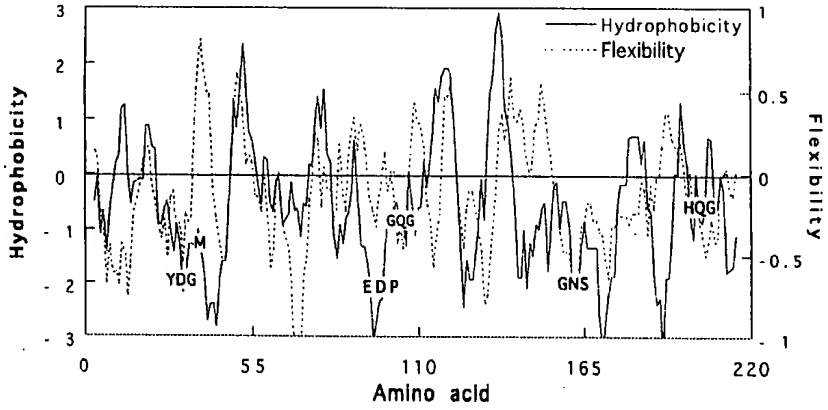
SEQUENCE

DIQLTQSPSSLSASVGDRVTITCRASQSVDYDGDSYMNWYQQKPGKAPKLLIYAAS-
YLESGVPSRFSGSGSGTDFTLTISSLQPEDFATYYCQQSHEDPYTFGQGTKVEIKRTV-
AAPSVFIFPPSDEQLKSGTASVVCLLNNFYPREAKVQWKVDNALQSGNSQESVTEQ-
DSKDSTYSLSSTLTLSKADYEKHKVYACEVTHQGLSSSPVTKSFNRGEC

REACTIVE SITES

.N.(6)	.D.(12)	.M.(1)	.Q.(15)
38 MNW	17 GDR	37 YMN	3 IQL
141 LNN	30 VDY		6 TQS
142 NNF	32 YDG		27 SQS
156 DNA	34 GDS		41 YQQ
162 GNS	74 TDF		42 QQK
214 FNR	86 EDF		83 LQP
	98 EDP		93 CQQ
	126 SDE		94 QQS
	155 VDN		104 GQG
	171 QDS		128 EQL
	174 KDS		151 VQW
	189 ADY		159 LQS
			164 SQE
			170 EQD
			203 HQG

HYDROFLEX PLOT



PREDICTED REACTIVITY AND DEGRADATION OF ANTIBODY E25 LIGHT CHAIN

The E25 antibody is a humanized monoclonal antibody that binds to human IgE and is under development for the treatment of asthma and other allergic diseases. The light chain has several reactive sites, including Asn-Gly, Asp-Gly, Asp-Pro, a single Met and Gln-Gly that may show chemical instability in aqueous solution. Recent studies have demonstrated the lability of the Asp-32 (in the YDG motif) towards isomerization, forming both cyclic imide and iso-Asp variants upon storage at pH 5.2 at room temperature (Cacia *et al.*, 1996). The Asp-32 residue also converts to the iso-Asp residue upon storage at pH 7.2 at room temperature, presumably through a cyclic imide intermediate. Both iso-Asp-32 and the cyclic imide variants show reduced binding to IgE. No other significant degradation products have been detected.

- **Antibody E25 Heavy Chain (451 residues)**

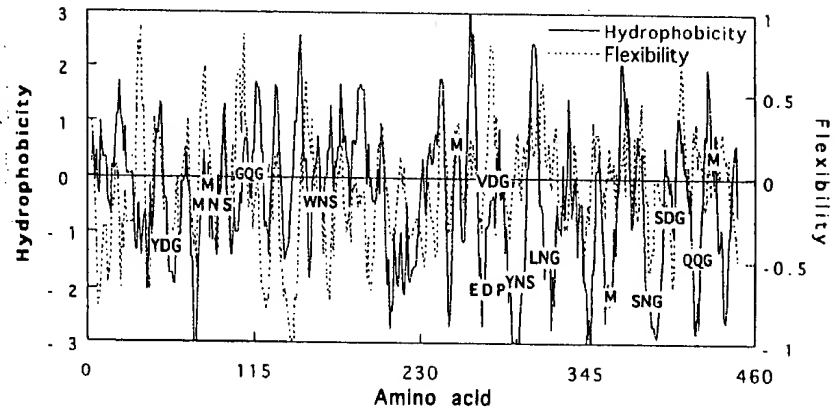
SEQUENCE

EVQLVESGGGLVQPGGSLRLSCAVSGYSITSGYSWNWIRQAPGKGLEWVASITYDG-
STNYNPSVKGRITISRDDSKNTFYLQMNSLRAEDTAVYYCARGSHYFGHWHFAVW-
GQGTLVTVSSASTKGPSVFPLAPSSKSTSGGTAALGCLVKDYFPEPVTVSWNSGALT-
SGVHTFPAVLQSSGLYSLSSVVTVPSSSLGTQTYICNVNHKPSNTKVDKKVEPKSCD-
KTHTCPPCPAPELLGGPSVFLFPPKPKDTLMISRTPEVTCVVVDVSHEDPEVKFNWY-
VDGVEVHNAKTKPREEQYNSTYRVVSVLTVLHQDWLNGKEYKCKVSNKALPA-
PIEKTISKAKGQPREPQVYTLPPSREEMTKNQVSLTCLVKGFYPSDIAVEWESNGQPE-
NNYKTTPPVLDSDGSFFLYSKLTVDKSRWQQGNVFSCSVMHEALHNHYTQKSL-
SLSPGK

REACTIVE SITES

.N.(20)	.D.(16)	.M.(4)	.Q.(16)
36 WNW	55 YDG	83 QMN	3 VQL
59 TNY	73 RDD	256 LMI	13 VQP
61 YNP	74 DDS	362 EMT	40 RQA
77 KNT	90 EDT	432 VMH	82 LQM
84 MNS	152 KDY		113 GQG
163 WNS	216 VDK		179 LQS
205 CNV	225 CDK		200 TQT
207 VNH	253 KDT		299 EQY
212 SNT	269 VDV		315 HQD
280 FNW	274 EDP		346 GQP
290 HNA	284 VDG		351 PQV
301 YNS	316 QDW		366 NQV
319 LNG	380 SDI		390 GQP
329 SNK	403 LDS		422 WQQ
365 KNQ	405 SDG		423 QQG
388 SNG	417 VDK		442 TQK
393 ENN			
394 NNY			
425 GNV			
438 HNH			

HYDROFLEX PLOT



PREDICTED REACTIVITY AND DEGRADATION OF ANTIBODY E25 HEAVY CHAIN

The E25 antibody is an anti-IgE antibody under development for the treatment of asthma and other allergic diseases (Presta *et al.*, 1993). This protein has several reactive sites, including

Asn-Gly, Asn-Ser, Asp-Gly, Asp-Pro, Met and Gln-Gly that may show chemical instability in aqueous solution. Recent studies have shown no significant degradation under mild conditions. In particular, Asp-55 (in the YDG motif) did not show evidence of isomerization, even though a similar Asp-Gly motif in the E25 light chain did show isomerization (see E25 light chain) (Cacia *et al.*, 1996). This variation in reactivity for Asp-Gly further illustrates the dependence of reactivity on tertiary structure, as well as on sequence motifs (Kossiakoff *et al.*, 1988).

- **Antibody Light Chain-κ (mouse) (214 residues)**

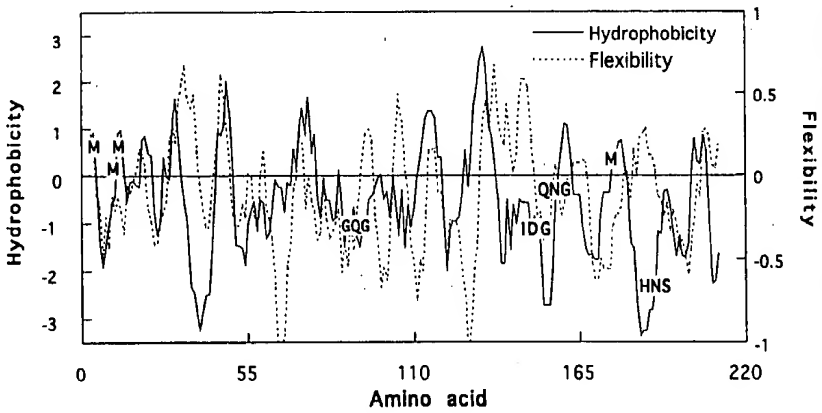
SEQUENCE

NIVMTQSPKSMSMSVGERVTLTCKASENVVTYVSWYQQKPEQSPKLLIYGASNRYT-
GVPDRFTGSGSATDFTLTISSVQAEDLADTHCGQGYSYPYTFGGGTKLEIKRADAAP-
TVSIFPPSSEQLTSGGASVVCFLNNFYPKDINVKWKIDGSERQNGVLBSBTXWBSKD-
STTSMSSTLTLTKDEYERHNSYTCEATHKTSTSPIVKSFNRNEC

REACTIVE SITES

.N.(12)		.D.(12)		.M.(4)	.Q.(8)
28 ENV	167 WBS	60 PDR	163 SBT	4 VMT	6 TQS
53 SNR	190 HNS	70 TDF	167 WBS	11 SMS	37 YQQ
137 LNN	210 FNR	82 EDL	170 KDS	13 SMS	38 QQK
138 NNF	212 RNE	85 ADT	184 KDE	175 SMS	42 EQS
145 INV		110 ADA			79 VQA
157 QNG		143 KDI			90 GQG
161 LBS		151 IDG			124 EQL
163 SBT		161 LBS			156 RQN

HYDROFLEX PLOT



PREDICTED REACTIVITY AND DEGRADATION OF KAPPA LIGHT CHAIN

The hydroflex plot for kappa light chain shows that there are only a few predicted reactive sites for degradation. These include Asn-157 within the -QNG- motif and several Met amino acids. Isolation and complete sequencing of the mouse kappa light chain was carried out over two decades ago, where it was found that the isolated product showed some micro-heterogeneity, likely due to deamidation at Asn-157 (Svanti and Milstein, 1972). The paper chromatography methods used make the assignment of this Asn rather ambiguous but plausible, considering the lack of other hot spots for deamidation in the same tryptic peptide.

Antibody OKT3 Heavy Chain (449 residues)

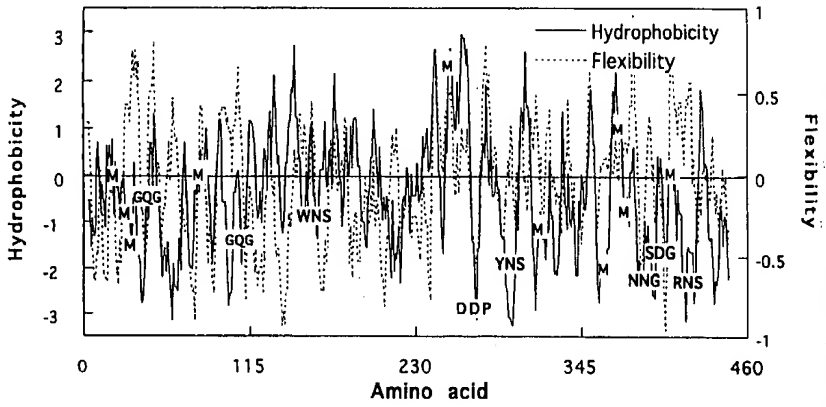
SEQUENCE

QVQLQQSGAELARPGASVKMSCKASGYTFTRYTMHWVKQRPGQGLEWIGYINPSR-
GYTNYNQKFKDKATLTTDKSSSTAYMQLSSLTSEDSAVYYCARYYDDHYCLDYWG-
QGTTLTVSSAKTTAPSVYPLAPVCGDTTGSSVTLGCLVKGYFPEPVTLTWNSGSLSS-
GVHTFPAVLQSDLYTLSSSVTVTSSTWPSQSITCNVAHPASSTKVDKKIEPRGPTIKPC-
PPCKCPAPNLLGGPSVFIFPPKIKDVLMISLSPIVTCVVDVSEDDPDVQISWFVNNVE-
VHTAQVQTHREDYNSTLRVVSALPIQHQDWMSGKEFKCKVNNKDLPAPIERTISKP-
KGSVRAPQVYVLPPPEEEMTKKQVTLTCMVTDFMPEDIYVEWTNNGKTELNYKNT-
EPVLDSDGSYFMYSKLRVEKKNWVERNSYSCSVVHEGLHNHHTTKSFSRTPGK

REACTIVE SITES

.N.(18)	.D.(21)		.M.(9)	.Q.(17)
52 INP	66 KDK	378 EDI	20 KMS	3 VQL
59 TNY	73 TDK	401 LDS	34 TMH	5 LQQ
61 YNQ	90 EDS	403 SDG	81 YMQ	6 QQS
161 WNS	101 YDD		254 LMI	39 KQR
202 CNV	102 DDH		316 WMS	43 GQG
235 PNL	107 LDY		360 EMT	62 NQK
282 VNN	136 GDT		370 CMV	82 MQL
283 NNV	179 SDL		375 FMP	111 GQG
299 YNS	213 VDK		408 FMY	177 LQS
326 VNN	251 KDV			197 SQS
327 NNK	267 VDV			276 VQI
385 TNN	271 EDD			290 AQT
386 NNG	272 DDP			292 TQT
392 LNY	274 PDV			311 IQH
395 KNT	297 EDY			313 HQD
418 KNW	314 QDW			349 PQV
423 RNS	329 KDL			364 KQV
436 HNH	373 TDF			

HYDROFLEX PLOT



PREDICTED REACTIVITY AND DEGRADATION OF OKT3 HEAVY CHAIN

The larger size of the OKT3 heavy chain makes it more likely that this is the more reactive chain, especially considering the large number of moderately reactive hot spots. This protein is predicted to react predominantly at Asn-386 (-NNG-) and possibly at some of the less reactive Asn-Ser sites. Further, this heavy chain has numerous Met residues, and so some oxidation might be expected. The major degradation pathway for this protein (as a part of the entire OKT3 complex) is at Asn-386 as predicted (Kroon *et al.*, 1992). Additionally, oxidation was observed at Met-34, Met-316, Met-360, and Met-408, most of which are found in fairly hydrophilic regions as predicted by hydropathy analysis. Sufficient oxidation of Met-34 was observed that the first OKT3 product formulation was eventually reformulated to include an inert headspace to reduce oxidation. A minor amount of deamidation was also found at Asn-423 (-RNS-), which is in a hydrophilic region of poor flexibility.

- **Antibody OKT3 Light Chain (213 residues)**

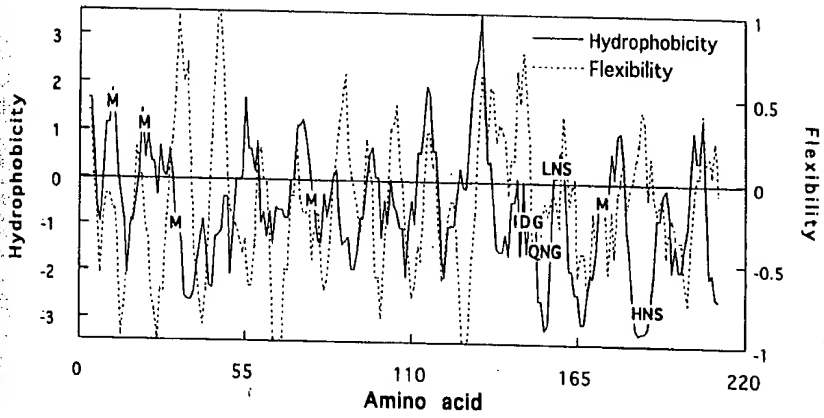
SEQUENCE

QIVLTQSPAIMSASPGEKVTMTCSASSSVSYMNWYQQKSGTSPKRWIYDTSKLASG-
VPAHFRGSGSGTSYSLTISGMEAEDAATYYCQQWSSNPFTFGSGTKLEINRADTAPT-
VSIFPPSSEQLTSGGASVVCFLNNFYPKDINVKWKIDGSERQNGVLNSWTDQDSKDS-
TYSMSSTLTLTKDEYERHNSYTCEATHKTSTSPIVKSFNRNEC

REACTIVE SITES

.N.(11)	.D.(9)	.M.(5)	.Q.(8)
33 MNW	49 YDT	11 IMS	6 RQS
93 SNP	81 EDA	21 TMT	36 YQQ
106 INR	109 ADT	32 YMN	37 QQK
136 LNN	142 KDI	77 GME	88 CQQ
137 NNF	150 IDG	174 SMS	89 QQW
144 INV	164 TDQ		123 EQL
156 QNG	166 QDS		155 RQN
160 LNS	169 KDS		165 DQD
189 HNS	183 KDE		
209 FNR			
211 RNE			

HYDROFLEX PLOT



PREDICTED REACTIVITY AND DEGRADATION OF OKT3 LIGHT CHAIN

The OKT3 antibody is a murine IgG2a antibody capable of binding CD3 and is used to clinically reverse rejections of human kidney transplants. Inspection of the hydroflex plot shows that the OKT3 light chain has several hot spots, of which the predominant site is predicted to be Asn-156, possibly followed by Asn-189, found in a hydrophilic region of predicted poor flexibility. The major degradation pathway for this protein (as a part of the entire OKT3 complex) in pH 7 PBS was at Asn-156 as predicted. A small amount of oxidative degradation occurred at Met-174, found in a region of intermediate hydropathy and flexibility. No other significant degradation was observed for the other potential hot spots (Kroon *et al.*, 1992).

Antibody OKT4a Heavy Chain (humanized) (447 residues)

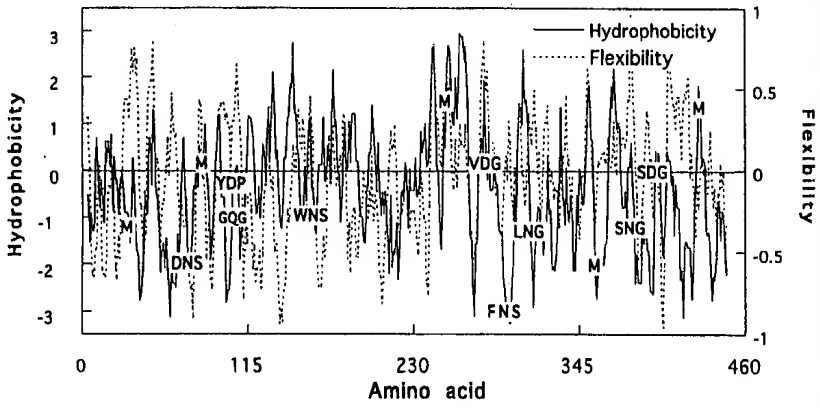
SEQUENCE

QVQLVESGGGVVQPGRSLRLSCSASGFTFSNYAMSWVRQAPGKGLEWVAAISDHST-
NTYYPDSVKGRFTISRDNSKNTLFLQMDSLRPEDTGVYFCARKYGGDYDPFDYWG-
QGTPVTVSSASTKGPSVFPLAPCSRSTSESTAALGCLVKDYFPEPVTVSWNSGALTSG-
VHTFPAVLQSSGLYSLSSVVTVPSSSLGTKTYTCNVDHKPSNTKVDKRVESKYGPP-
CPSCPAPEFLGGPSVFLFPPKPKDTLMISRTPEVTCVVDVSQEDPEVQFNWYVDGVE-
VHNAKTKPREEQFNSTYRVVSVLTVLHQDWLNGKEYKCKVSNKGPSSIEKTISKAK-
GQPREPQVYTLLPPSQEEMTKNQVSLTCLVKGFYPSDIAVEWESNGQPENNYKTTP-
PVLDSDGSFFLYSRLTVDKSRWQEGNVFSCSVMHEALHNHYTQKSLSLSLGK

REACTIVE SITES

.N.(18)				.D.(20)				.M.(5)		.Q.(17)	
31	SNY	315	LNG	53	SDH	215	VDK	34	AMS	3	VQL	342	GQP
57	TNT	325	SNK	62	PDS	249	KDT	83	QMD	13	VQP	347	PQV
74	DNS	361	KNQ	73	RDN	265	VDV	252	LMI	39	RQA	355	SQE
77	KNT	384	SNG	84	MDS	270	EDP	358	EMT	82	LQM	362	NQV
162	WNS	389	ENN	90	EDT	280	VDG	428	VMH	112	GQG	386	GQP
204	CNV	390	NNY	103	GDY	312	QDW			178	LQS	418	WQE
211	SNT	421	GNV	105	YDP	376	SDI			268	SQE	438	TQK
276	FNW	434	HNH	108	FDY	399	LDS			274	VQF		
286	HNA			151	KDY	401	SDG			295	EQF		
297	FNS			206	VDH	413	VDK			311	HQD		

HYDROFLEX PLOT



PREDICTED REACTIVITY AND DEGRADATION OF OKT4a HEAVY CHAIN

The larger size of the OKT4a heavy chain makes it the likely reactive chain, especially considering the large number of moderately reactive hot spots. This protein is predicted to react predominantly at Asn-315 (-LNG-) and Asn-384 (-SNG-), and possibly at some of the less reactive Asn-Ser sites. Reaction is also predicted at Asp-Gly to give iso-Asp-Gly (although this is often difficult to detect experimentally), as well as at Asp-Pro at lower pHs. This heavy chain has numerous Met residues, and thus some oxidation might be expected. The major degradation pathway for OKT4 heavy chain (as a part of the entire OKT4a complex) at pH's less than 6.5 was cleavage at Asp-270 within the -EDP- motif (Kroon, 1994). Interestingly, no cleavage was found at Asp-105 within the -YDF- motif. A minor amount of cleavage was observed at bonds N-terminal to several Ser and Thr residues, including Ser-220, Thr-250, Thr-335, and Thr-350. Deamidation was found to be slow for this protein below neutral pH; however, the exact sites of deamidation were not determined, and deamidation was identified only by an acidic shift in the IEF pattern.

Antibody OKT4a Light Chain (humanized) (214 residues)

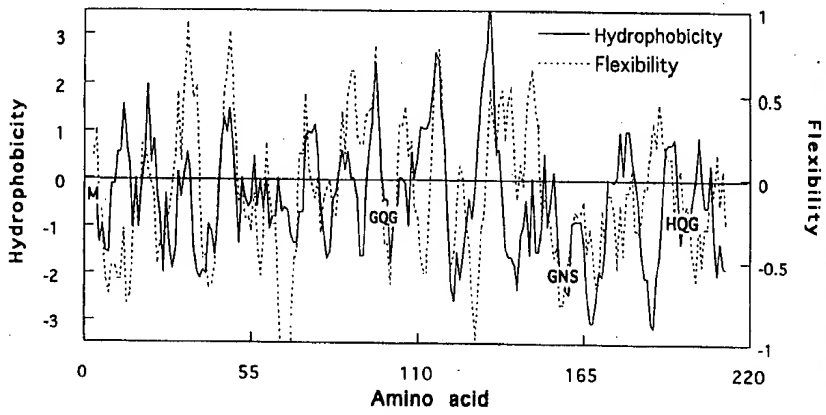
SEQUENCE

DIQMTQSPSSLSASVGDRVTITCKASQDINNYIAWYQQTPGKAPKLLIHYTSTLQPG-
VPSRFSGSGSGTDYTFTISSLQPEDIATYYCLQYDNLLFTFGQGTKLQITRTVAAPSVF-
IFPPSDEQLKSGTASVVCLLNNFYPREAKVQWKVDNALQSGNSQESVTEQDSKDST-
YSLSSTLTLSKADYEKHKVYACEVTHQGLSSPVTKSFNRGEC

REACTIVE SITES

.N.(8)		.D.(10)		.M.(1)		.Q.(16)	
30	INN	17	GDR	4	QMT	3	IQM
31	NNY	28	QDI			6	TQS
93	DNL	70	TDY			27	SQD
137	LNN	82	EDI			37	YQQ
138	NNF	92	YDN			38	QQT
152	DNA	122	SDE			55	LQP
158	GNS	151	VDN			79	LQP
210	FNR	167	QDS			90	LQY
		170	KDS			100	GQG
		185	ADY			105	LQI
						124	EQL
						147	VQW
						155	LQS
						160	SQE
						166	EQD
						199	HQG

HYDROFLEX PLOT



PREDICTED REACTIVITY AND DEGRADATION OF OKT4a LIGHT CHAIN

This light chain has few reactive sites, suggesting that the OKT4a heavy chain is the major site of chemical degradation. A minor amount of cleavage at Ser-203 was found as a trace reaction. No oxidation of Met-4 was reported (Kroon, 1994).

- **Atrial Natriuretic Peptide (ANP) (human) (28 residues)**

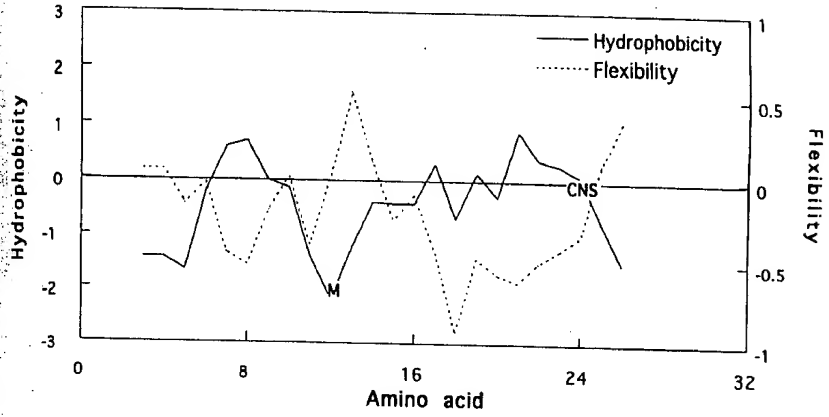
SEQUENCE

SLRRSSCFGGRMDRIGAQSGLGCNSFRY

REACTIVE SITES

.N.(1)	.D.(1)	.M.(1)	.Q.(1)
24 CNS	13 MDR	12 RMD	18 AQS

HYDROFLEX PLOT



PREDICTED REACTIVITY AND DEGRADATION OF ANP

The primary amino acid sequence for ANP shows that this peptide has only one of the traditional hydrolytic hot spots, Asn-Ser, and lacks the Asn-Gly, Asp-Gly, and Gln-Gly hot spots. Asn-24 resides within the -CNS- motif and is expected to be reactive based on its primary amino acid sequence. It does have, however, a single Met that is capable of being oxidized. Two degradation pathways have been observed for this cyclic peptide, deamidation of Asn-24 and oxidation of Met-12 (Wang, 1995).

Brain-Derived Neurotrophic Factor (BDNF) (human) (120 residues)

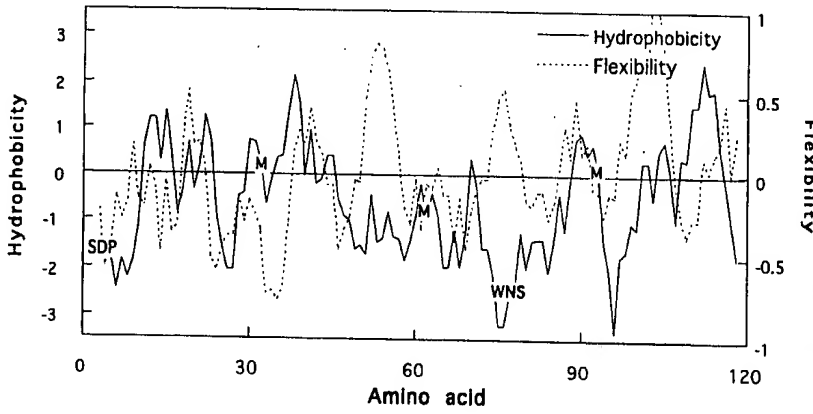
SEQUENCE

MHSDPARRGELSVCDSISEWVTAADKKTAVDMSGGTVTVLEKVPVSKGQLKQYFY-ETKCNPMGYTKEGCRGIDKRHWNSQCRTTQSYVRALTMDSKKRIGWRFIRIDTSCV-CTLTIKRGR

REACTIVE SITES

.N.(2)	.D.(7)	.M.(4)	.Q.(4)
60 CNP	4 SDP	73 IDK	1 MH	49 GQL
78 WNS	15 CDS	94 MDS	32 DMS	52 KQY
	25 ADK	107 IDT	62 PMG	80 SQC
	31 VDM		93 TMD	85 TQS

HYDROFLEX PLOT



PREDICTED REACTIVITY AND DEGRADATION OF BDNF

This protein is relatively free of activated hot spots, except for Met and the acid-sensitive Asp-Pro motive. Some degradation studies have been carried out in the neutral pH range, where it was found that the primary degradation pathways were cleavage at His-2–Ser-3 and between Asp-4–Pro-5 (Hershensen *et al.*, 1995). Oxidation at Met-1 and Met-62 was also observed, with minor amounts of oxidation at Met-32. Other minor degradation pathways included cleavage at Val-45–Ser-46, Lys-47–Gly-48, and Asn-60–Pro-61 (reaction conditions not specified).

- **Calbindin (bovine) (76 residues)**

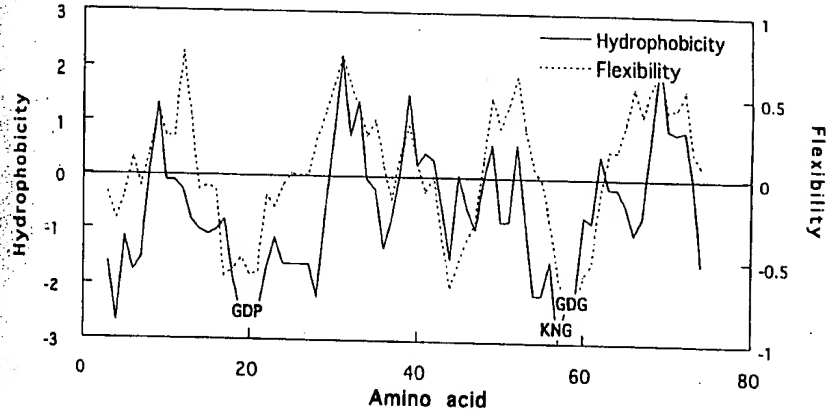
SEQUENCE

MKSPEELKGIFEKYAAKEGDPNQLSKEELKLLLQTEFPSLLKGPSTLDELFEELDKN-GDGEVSFEEFQVLVKKISQ

REACTIVE SITES

.N.(2)	.D.(4)	.M.(0)	.Q.(4)
22 PNQ	20 GDP		23 NQL
57 KNG	48 LDE		34 LQT
	55 LDK		68 FQV
	59 GDG		76 SQ

HYDROFLEX PLOT



PREDICTED REACTIVITY AND DEGRADATION OF CALBINDIN

The hydroflex plot for calbindin shows that this protein contains the reactive Asn-Gly hot spot within a region that is predicted to be fairly hydrophilic and flexible. An Asp-Gly motive is also found nearby in this hydrophilic region. Preparations of recombinant bovine calbindin D9k have been shown to be heterogeneous by IEF, due to deamidation of Asn-57 within the -KNG- motif (Chazin *et al.*, 1989). Calbindin also contains an Asp-Gly in the same region, but no degradation at the Asp-Gly site or at the acid-labile Asp-Pro site was reported.

Calmodulin (148 residues)

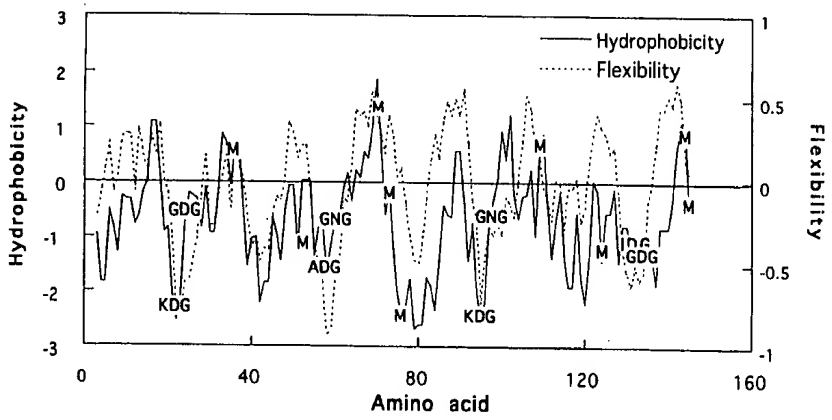
SEQUENCE

ADQLTEEQIAEFKEAFSLFDKDGDGTITTKELGTVMRSLGQNPTEAELQDMINEVDA-DGNGTIDFPEFLTMMARKMKDTDSEEEIREAFRVFDKDGNGYISAAELRHVMTNLG-EKLTDEEVDEMIREADIDGDGQVNYEEFVQMMTAK

REACTIVE SITES

.N.(6)	.D.(17)		.M.(9)	.Q.(6)
42 QNP	2 ADQ	80 TDS	36 VMR	3 DQL
53 INE	20 FDK	93 FDK	51 DMI	8 EQI
60 GNG	22 KDG	95 KDG	71 TMM	41 GQN
97 GNG	24 GDG	118 TDE	72 MMA	49 LQD
111 TNL	50 QDM	122 VDE	76 KMK	135 GQV
137 VNY	56 VDA	129 ADI	109 VMT	143 VQM
	58 ADG	131 IDG	124 EMI	
	64 IDF	133 GDG	144 QMM	
	78 KDT		145 MMT	

HYDROFLEX PLOT



PREDICTED REACTIVITY AND DEGRADATION OF CALMODULIN

Calmodulin contains at least eight sites that may undergo deamidation or cyclization, as well as numerous Met residues. All of the -XDG- and -XNG- reactive motifs lie in moderately hydrophilic regions of good predicted flexibility, further supporting the notion that calmodulin should be particularly susceptible to hydrolytic degradation. Calmodulin has two Asn-Gly sites, which are predicted to be more reactive than the Asp-Gly sites. Measurements of ammonia release and methyl transfer rates showed that calmodulin was extremely reactive towards hydrolytic degradation, giving 0.5 mole of ammonia released per mole calmodulin at pH 7.4 and 37°C after 8–9 days (Johnson *et al.*, 1989a). Comparison measurements of ammonia release and methyl transfer with other proteins showed that calmodulin is much more reactive than the other proteins surveyed. Although the entire degradation profile for calmodulin was not determined, it was believed that the primary sites of deamidation were Asn-60 (-GNG-) and Asn-97 (-GNG-). Calmodulin has numerous methionine residues, and the C-terminal residues are most susceptible to oxidation by peroxynitrite (Hühmer *et al.*, 1996) or by hydrogen peroxide (Yao *et al.*, 1996).

• Carbonic Anhydrase C (259 residues)

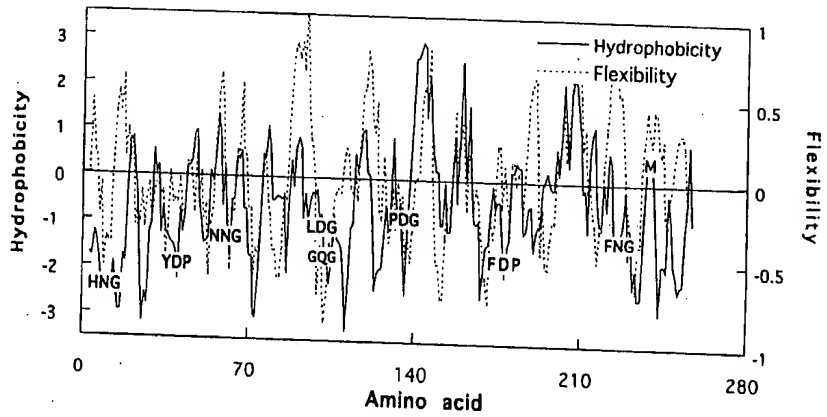
SEQUENCE

SHHWGYGKHNGPEHWHKDFPIAKGERQSPVDIDTHTAKYDPSLKPLSVSYDQATSL-
RILNNGHAFNVEFDDSEDKAVLKGGPLDGTYRLIQFHFHWGSLDGQGSQHTVDKK-
KYAAELHLVHWNTKYGDFGKAVQQPDGLAVLGIFLKVGSAKPGLQKVVDVLDSIK-
TKGKSADFTNFDPRGLLPESLDYWTYPGSLTTPPLLECVTWIVLKEPISVSSEQVLKF-
RKLNFNGEGEPEELMVDNWRPAQPLKNRQIKASFK

REACTIVE SITES

.N.(10)	.D.(19)	.M.(1)	.Q.(11)
10 HNG	18 KDF	239 LMV	27 RQS
60 LNN	31 VDI		52 DQA
61 NNG	33 IDT		91 IQF
66 FNV	40 YDP		102 GQG
123 WNT	51 YDQ		105 SQH
176 TNF	70 FDD		134 VQQ
228 LNF	71 DDS		135 QQP
230 FNG	74 EDK		156 LQK
242 DNW	84 LDG		220 EQV
251 KNR	100 LDG		247 AQP
	109 VDK		253 RQI
	128 GDF		
	137 PDG		
	160 VDV		
	163 LDS		
	173 ADF		
	178 FDP		
	188 LDY		
	241 VDN		

HYDROFLEX PLOT



PREDICTED REACTIVITY AND DEGRADATION OF CARBONIC ANHYDRASE C

This protein has several residues predicted to be reactive, including three Asn-Gly motifs. In one of the original papers describing the primary structure of human carbonic anhydrase C it was noted during the sequence analysis work that several residues underwent facile deamida-

tion. All were identified as Asn-Gly sequences (Henderson *et al.*, 1976). It was not determined if the protein was deamidated before isolation, during its purification, or during peptide analysis. Several of the steps used were carried out at elevated temperatures or used strong acid (1 M acetic acid for example), and so it was not possible to determine the origin of the protein microheterogeneity.

CD4 (human) (370 residues)

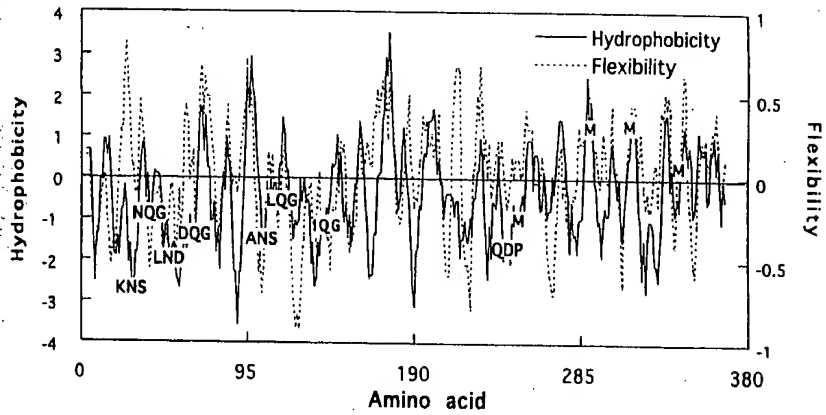
SEQUENCE

KKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIKILGNQGSFLTKGPSKLNDRAD-
SRRSLWDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQLLVFGLTANSDTHLLQGQSL-
TLTLESPPGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSGTWTCTVLQNQKKVEFK-
IDIVVLAFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWWQAERASSSKSWITF-
DLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLALEAKTGKLH-
QEVNLVVMRATQLQKNLTCEVWGPTSPKLMLSLKLENKEAKVSKREKAVWVLNPE-
AGMWQCLLSDSGQVLLESNIKVLPTWSTPVH

REACTIVE SITES

.N.(16)	.D.(13)	.M.(4)	.Q.(27)
30 KNS	10 GDT	249 QMG	20 SQK
32 SNQ	53 NDR	292 VMR	25 IQF
39 GNQ	56 ADS	314 LML	33 NQI
52 LND	63 WDQ	342 GMW	40 NQG
66 GNF	78 EDS		64 DQG
73 KNL	80 SDT		89 DQK
103 ANS	88 EDQ		94 VQL
137 KNI	105 SDT		110 LQG
164 QNQ	153 QDS		112 GQS
233 KNK	173 IDI		129 VQC
271 GNL	230 FDL		139 IQG
288 VNL	244 QDP		148 SQL
300 KNL	349 SDS		152 LQD
321 ENK			163 LQN
337 LNP			165 NQK
358 SNI			180 FQK
			193 EQV
			216 WQA
			243 TQD
			248 LQM
			261 PQA
			265 PQY
			285 HQE
			296 TQL
			298 LQK
			344 WQC
			352 GQV

HYDROFLEX PLOT



PREDICTED REACTIVITY AND DEGRADATION OF CD4

This protein harbors four Gln-Gly sites and two Asn-Ser sites, all of which are commonly regarded as the potential hot spots for degradation. CD4 also has four Met residues in the C-terminal end of the molecule. An elegant study on the deamidation of soluble CD4 has been reported, wherein it was found that Asn-52 (in the -LND- motif) was the primary degradation site at pH 7.2 and 25°C (Teshima *et al.*, 1991a, 1995a). The -LND- motif is generally thought to be fairly unreactive, and so this is a clear-cut example where deamidation may occur in aqueous formulations at sites other than Asn-Gly or Asn-Ser. It is interesting to note that Asn-52 resides in a region predicted to be moderately hydrophilic and flexible, in good agreement with its crystal structure of the V1 and V2 domains (Wang *et al.*, 1990) and this may contribute to its reactivity. No oxidation at Met was observed.

CD4-IgG (407 residues)

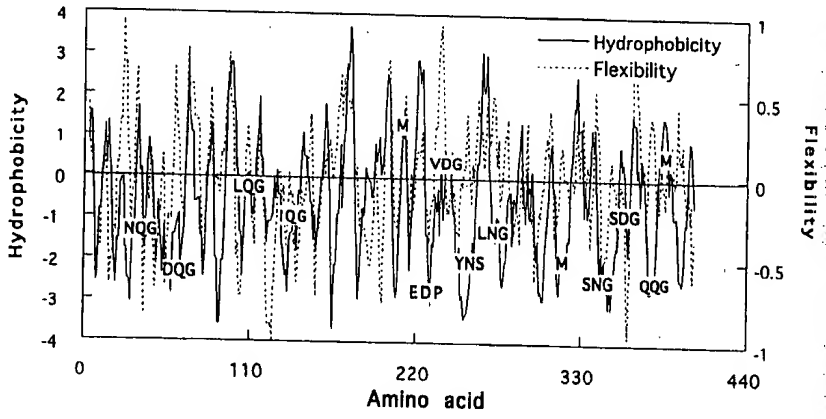
SEQUENCE

KKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIKILGNQGSFLTKGPSKLNDRAD-
SRRSLWDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQLLVFGLTANSDTHLLQGQSL-
TLTLESPPGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSGTWTCTVLQNQKKVEFK-
IDIVVLAFQDKTHTCPPCPAPELLGGPSVFLFPPKPKDTLMISRTPEVTCVVVDVSHED-
PEVKFNWYVDGVEVHNAKTKPREEQYNSTYRVVSVLTVLHQDWLNGKEYKCKVS-
NKALPAPIEKTISKAKGQPREPQVYTLPPSREEMTKNQVSLTCLVKGFYPSDIAVEWE-
SNGQPENNYKTTPPVLDSDGSFFLYSKLTVDKSRWQQGNVFSCSVMHEALHNHYT-
QKSLSLSPGK

REACTIVE SITES

.N.(20)	.D.(20)	.M.(3)	.Q.(25)	
30 KNS	10 GDT	212 LMI	20 SQK	322 NQV
32 SNQ	53 NDR	318 EMT	25 IQF	346 GQP
39 GNQ	56 ADS	388 VMH	33 NQI	378 WQQ
52 LND	63 WDQ		40 NQG	379 QQG
66 GNF	78 EDS		64 DQG	398 TQK
73 KNL	80 SDT		89 DQK	
103 ANS	88 EDQ		94 VQL	
137 KNI	105 SDT		110 LQG	
164 QNQ	153 QDS		112 GQS	
236 FNW	173 IDI		129 VQC	
246 HNA	181 QDK		139 IQG	
257 YNS	209 KDT		148 SQL	
275 LNG	225 VDV		152 LQD	
285 SNK	230 EDP		163 LQN	
321 KNQ	240 VDG		165 NQK	
344 SNG	272 QDW		180 FQD	
349 ENN	336 SDI		255 EQY	
350 NNY	259 LDS		271 HQD	
381 GNV	361 SDG		302 GQP	
394 HNH	373 VDK		307 PQV	

HYDROFLEX PLOT



PREDICTED REACTIVITY AND DEGRADATION OF CD4-IgG

This protein has partial identity with CD4 (Harris *et al.*, 1990) and so might be expected to degrade at the same hot spots in this region. This CD4 molecule also has two Asn-Gly and two

Asp-Gly moieties (the two most reactive hot spots) and so is predicted to degrade at these hot spots. The observed degradation pathway of CD4-IgG was found to be similar to CD4, that is, degradation at Asn-52 (in the -LND- motif) (Teshima and Yim, 1995b; Teshima and Wu, 1996). Again, the -LND- motif is generally thought to be fairly unreactive, and so this example illustrates that deamidation may occur in aqueous formulations at sites other than Asn-Gly or Asn-Ser.

CD4-PE40 (545 residues)

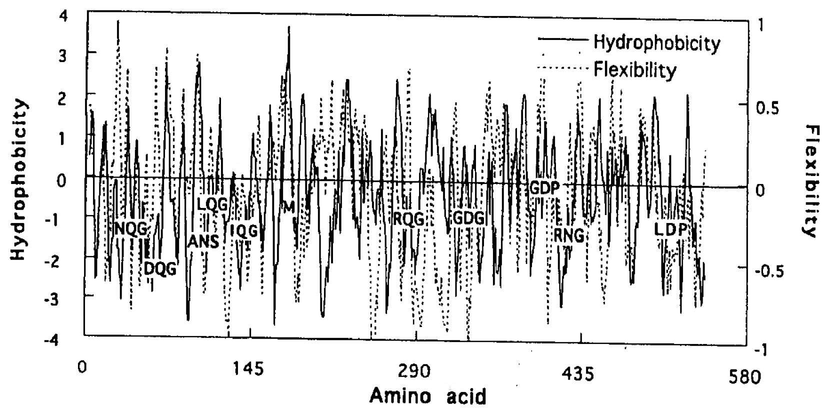
SEQUENCE

MKKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIKILGNQGSFLTKGPSKLNDRA-
DSRRSLWDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQLLVFGLTANSDTHLLQGQS-
LTLTLESPPGSSPVQCRSPRGKNIQGGKTLSVSQLELQDSGTWTCTVLQNQKKVEFK-
IDIVVLAHMAEEGGSLAALTAHQACHLPLETFTRHRQPRGWEQLEQCGYPVQRLVA-
LYLAARLSWNQVDQVIRNALASPGSGGDLGEAIREQPEQARLALTLAAAESERFVR-
QGTGNDEAGAANADVVSLTCPVAAGECAGPADSGDALLERNYPTGAEFLGDGGDV-
SFSTRGTQNWTVERLLQAHRQLEERGYVFVGYHGTFLEAAQSIVFGGVRARSQDL-
DAIWRGFYIAGDPALAYGYAQDQEPDARGRIRNGALLRVYVPRSSLPGFYRTSLTLA-
APEAAGEVERLIGHPLPLRLDAITGPEEEGGRLETILGWPLAERTVVIPSAIPTDPRNV-
GGDLDPSSIPDKEQAISALPDYASQPGKPPREDLK

REACTIVE SITES

.N.(17)	.D.(30)		.M.(1)	.Q.(34)	
31 KNS	11 GDT	338 GDV	181 HMA	21 SQK	215 EQL
33 SNQ	54 NDR	393 QDL		26 IQF	218 EQC
40 GNQ	57 ADS	395 LDA		34 NQI	224 VQR
53 LND	64 WDQ	406 GDP		41 NQG	239 NQV
67 GNF	79 EDS	416 QDQ		65 DQG	242 DQV
74 KNL	81 SDT	420 PDA		90 DQK	264 EQP
104 ANS	89 EDQ	472 LDA		95 VQL	267 EQA
138 KNI	106 SDT	506 TDP		111 LQG	285 RQG
165 QNQ	154 QDS	513 GDL		113 GQS	347 TQN
238 WNQ	174 IDI	515 LDP		130 VQC	356 LQA
246 RNA	241 VDQ	521 PDK		140 IQG	360 RQL
289 GND	256 GDL	531 PDY		149 SQL	380 AQS
296 ANA	290 NDE	543 EDL		153 LQD	392 SQD
325 RNY	298 ADV			164 LQN	415 AQD
348 QNW	316 ADS			166 NQK	417 DQE
427 RNG	319 GDA			195 HQA	524 EQA
509 RNV	335 GDG			209 RQP	535 SQP

HYDROFLEX PLOT



PREDICTED REACTIVITY AND DEGRADATION OF CD4-PE40

The recombinant human CD4-*Pseudomonas* exotoxin hybrid protein shows selective killing of HIV-1 infected cells and thus represents a novel therapeutic agent for the treatment of AIDS (Chaudhary *et al.*, 1988). Comparative analysis of this protein with soluble CD4 (see previous entry) provides additional insight into protein degradation in aqueous solution. CD4-PE40 has a single Asn-Gly site which is predicted to be reactive, as well as a number of lesser reactive sites such as Gln-Gly. This protein also has a single Met near the conjugation site of CD4 and PE40. The major degradation site for this protein in aqueous solution was Met oxidation, with no other clearly detectable degradation pathways noted (Hageman, 1995). Soluble CD4 did not show degradation at Met, because soluble CD4 does not have a Met at this position (see comparison of the soluble CD4 and CD4-PE40 amino acid sequences in Scheme 9). Of note, however, was the lack of degradation at Asn-427 in the RNG motif in CD4-PE40 (a predicted hot spot); this may be due to the conformational nature of the protein about this motif. It is also of interest to note that CD4-PE40 did not show any deamidation at Asn-53 within the -LND- motif, observed to be the major site of degradation of soluble CD4 in aqueous solution (see previous entry). Because the CD4 binding activity of CD4-PE40 is similar to soluble CD4, one must assume that the conformation of the CD4 region in CD4-PE40 is similar to soluble CD4. Thus, this protein provides some contrast to the "unusual" degradation pathway for soluble CD4, in that no degradation was observed at the -LND- motif generally thought to be fairly unreactive. This is likely due to the different methods of analysis used.

```
                  10        20        30        40
CD4        KKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIKILGNQGSFLTKGPS
           *************************************************
CD4-PE40  MKKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIKILGNQGSFLTKGPS
                  10        20        30        40        50

          50        60        70        80        90
CD4       KLNDRADSRRSLWDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQLLVFG
          **************************************************
CD4-PE40  KLNDRADSRRSLWDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQLLVFG
                  60        70        80        90       100

          100       110       120       130       140
CD4       LTANSDTHLLQGQSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLSVSQL
          **************************************************
CD4-PE40  LTANSDTHLLQGQSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLSVSQL
                 110       120       130       140       150

          150       160       170       180       190
CD4       ELQDSGTWTCTVLQNQKKVEFKIDIVVLAFQKASSIVYKKEGEQVEFSFP
          ****************************           .    .   .*
CD4-PE40  ELQDSGTWTCTVLQNQKKVEFKIDIVVLAHMAEEGGSLAALTAHQACHLP
                 160       170       180       190       200

          200       210       220       230       240
CD4       L-AFTVEKLTGSGELWWQAERAS-SSKSWITFDLKNKEVSVKRVTQDPKL
          * .** ..       * * *. .  . ... *    *  . .* *  .
CD4-PE40  LETFTRHRQPRG---WEQLEQCGYPVQRLVALYLAAR-LSWNQVDQVIRN
                 210           220       230        240

          250       260       270       280       290
CD4       QMGKKLPLHLTLPQALPQYAGSGNLTLALEAKTGK--LHQEVNLVVMRAT
          ..  .   * .*. . .  . *.*.* *  .  ..*          *.
CD4-PE40  ALASPGS-GGDLGEAIREQPEQARLALTLAAAESERFVRQGTGNDEAGAA
            250       260       270       280       290

            300            310       320       330
CD4       QLQ-KNLTCEV-----WGPTSPKLMLSLKLENKEAKVSKREKAVWVLNPE
          . .  .*** *       **. .   *        *         *
CD4-PE40  NADVVSLTCPVAAGECAGPADSGDALLERNYPTGAEFLGDGGDVSFSTRG
            300       310       320       330       340

          340       350       360       370
CD4       AGMWQCLLSDSGQVLLESNIKVLPTWSTPVH
          . *    ..    **   *.
CD4-PE40  TQNWTVERLLQAHRQLEERGYVFVGYHGTFLEAAQSIVFGGVRARSQDLD
            350       360       370       380       390

CD4-PE40  AIWRGFYIAGDPALAYGYAQDQEPDARGRIRNGALLRVYVPRSSLPGFYR
            400       410       420       430       440

CD4-PE40  TSLTLAAPEAAGEVERLIGHPLPLRLDAITGPEEEGGRLETILGWPLAER
            450       460       470       480       490

CD4-PE40  TVVIPSAIPTDPRNVGGDLDPSSIPDKEQAISALPDYASQPGKPPREDLK
            500       510       520       530       540
```

Scheme 9. Sequence comparison of CD4 and CD4-PE40.

Chloroperoxidase (*Caldariomyces fumago*) (300 residues)

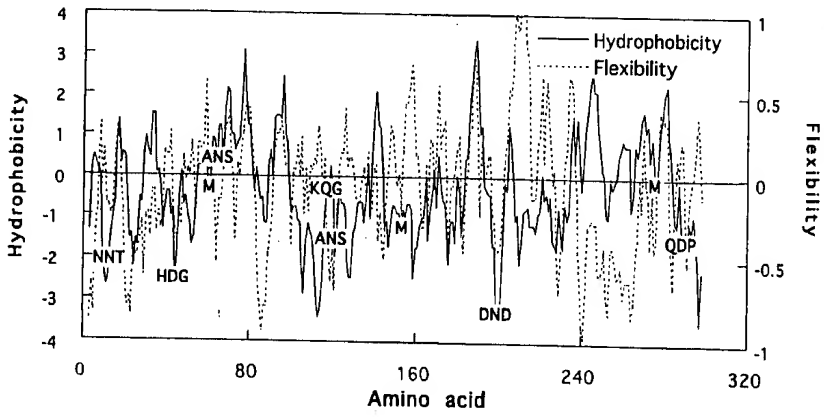
SEQUENCE

EPGSGIGYPYDNNTLPYVAPGPTDSRAPCPALNALANHGYIPHDGRAISRETLQNAF-
LNHMGIANSVIELALTNAFVVCEYVTGSDCGDSLVNLTLLAEPHAFEHDHSFSRKDY-
KQGVANSNDFIDNRNFDAETFQTSLDVVAGKTHFDYADMNEIRLQRESLSNELDFP-
GWFTESKPIQNVESGFIFALVSDFNLPDNDENPLVRIDWWKYWFTNESFPYHLGWH-
PPSPAREIEFVTSASSAVLAASVTSTPSSLPSGAIGPGAEAVPLSFASTMTPFLLATNAP-
YYAQDPTLRPQRQA

REACTIVE SITES

.N.(21)		.D.(19)		.M.(3)	.Q.(8)
12 DNN	129 RNF	11 YDN	152 ADM	61 HMG	54 LQN
13 NNT	154 MNE	24 TDS	168 LDF	153 DMN	116 KQG
33 LNA	165 SNE	44 HDG	193 SDF	276 TMT	136 FQT
37 ANH	181 QNV	86 SDC	198 PDN		159 LQR
55 QNA	195 FNL	89 GDS	200 NDE		180 IQN
59 LNH	199 DND	106 HDH	208 IDW		183 VQS
65 ANS	202 ENP	113 KDY	291 QDP		290 AQD
74 TNA	216 TNE	123 NDF			297 PQR
93 VNL	284 TNA	126 IDN			299 RQA
120 ANS		131 FDA			
122 SND		140 LDV			
127 DNR		149 FDY			

HYDROFLEX PLOT



PREDICTED REACTIVITY AND DEGRADATION OF CHLOROPEROXIDASE

This glycoprotein has several sites that may undergo degradation in aqueous solution, including pyroglutamic acid formation at the N-terminus. This was confirmed experimentally, in that approximately two thirds of the protein resisted Edman degradation, indicative of a blocked N-terminus. Purification of chloroperoxidase from the filamentous fungus *Caldariomyces fumago* showed microheterogeneity at Asn-13 (-NNT-), Asn-199 (-DND-), and Gln-183 (converted completely to -VES-) (Kenigsberg *et al.*, 1987), all at sites not thought to be traditional hot spots. Unfortunately the work-up of this protein had a heat-inactivation step (pH >8, 100°C for 2 min), which may account for some of the deamidation observed at these sites. Further, no controls were carried out to show that these chemical modifications were due to enzymatic hydrolysis during the lengthy work-up.

Cholera B Subunit Protein (*Vibrio cholerae*) (103 residues)

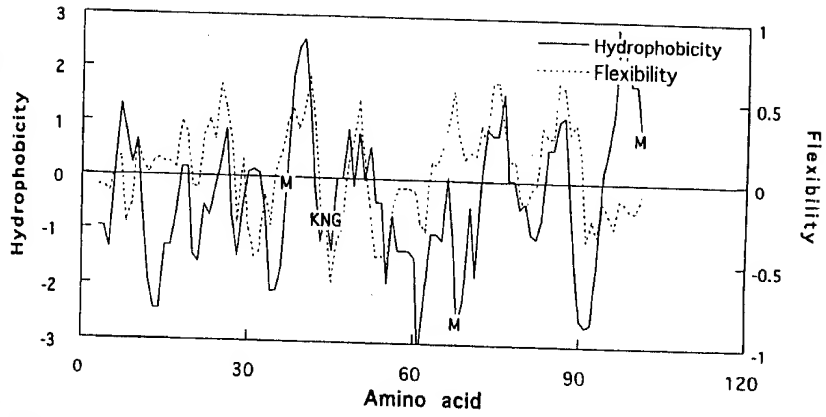
SEQUENCE

TPQNITDLCAEYHNTQIHTLNNKIFSYTESLAGKREMAIITFKNGATFEVEVPGSQHI-
DSQKKAIERMKNTLRIAYLTEAKVEKLCVWNNKTPHAIAAISMAN

REACTIVE SITES

.N.(9)		.D.(2)	.M.(3)	.Q.(4)
4 QNI	70 KNT	7 TDL	37 EMA	3 PQN
14 HNT	89 WNN	59 IDS	68 RMK	16 TQI
21 LNN	90 NNK		101 SMA	56 SQH
22 NNK	103 AN			61 SQK
44 KNG				

HYDROFLEX PLOT



PREDICTED REACTIVITY AND DEGRADATION OF CHOLERA B TOXIN

This subunit protein contains only two different residues predicted to be reactive: Asn-Gly (-KNG-) and Met. Mass spectral analysis and Edman degradation of peptides derived from the B subunit of *Vibrio cholerae* toxin showed microheterogeneity at Asn-44 found within the -KNG- motif (Takao *et al.*, 1985). This is the only site predicted to show hydrolytic reactivity based on primary amino acid sequence and hydroflex analysis. Interestingly, the authors reported Asp instead of Asn at position 22 (-NNK-) and position 70 (-KNT-), even though Asn was found at these positions in some of the earlier cholera B toxin strains (shown in the sequence above). However, the nucleotide sequences encode for both Asp and Asn (depending on the strain), and so Asp for Asn at these sites should not be considered conclusively as microheterogeneity. No indication of Met oxidation in this protein was reported.

Ciliary Neurotrophic Factor (CNTF) (human) (199 residues)

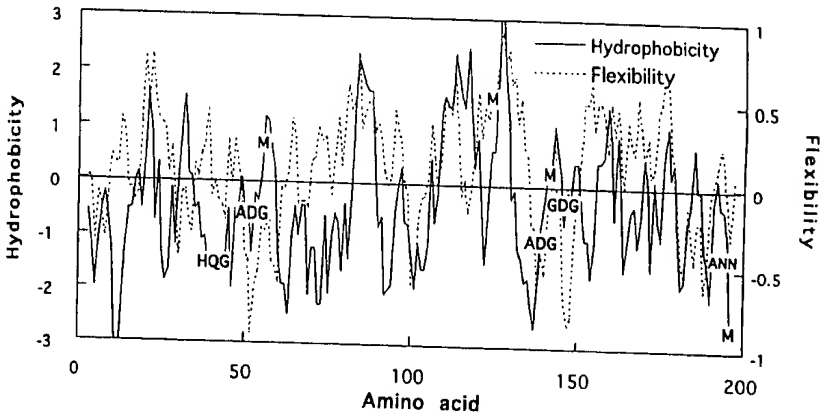
SEQUENCE

AFTEHSPLTPHRRDLCSRSIWLARKIRSDLTALTESYVKHQGLNKNINLDSADGMPV-
ASTDQWSELTEAERLQENLQAYRTFHVLLARLLEDQQVHFTPTEGDFHQAIHTLLL-
QVAAFAYQIEELMILLEYKIPRNEADGMPINVGDGGLFEKKLWGLKVLQELSQWTV-
RSIHDLRFISSHQTGIPARGSHYIANNKKM

REACTIVE SITES

.N.(8)	.D.(10)	.M.(4)	.Q.(12)
44 LNK	14 RDL	55 GMP	41 HQG
46 KNI	29 SDL	126 LMI	62 DQW
48 INL	50 LDS	141 GMP	73 LQE
75 ENL	53 ADG	199 KM	77 LQA
136 RNE	61 TDQ		93 DQQ
144 INV	92 EDQ		94 QQV
195 ANN	103 GDF		106 HQA
196 NNK	139 ADG		114 LQV
	147 GDG		121 YQI
	174 HDL		162 LQE
			166 SQW
			182 HQT

HYDROFLEX PLOT



PREDICTED REACTIVITY AND DEGRADATION OF CNTF

Inspection of the hydroflex plot for CNTF shows that this protein has a few moderately reactive hot spots: three Asp-Gly residues and four Mets. None of the Asp-Gly are found in highly hydrophilic regions, and so might be expected to show reduced reactivity (if any at all). The major degradation pathway for CNTF was recently deduced and found not to involve any of the traditional hot spots; deamidation was observed at Asn-195 in the -ANN- motif close to the C-terminus (Maneri, 1994). Although deamidation takes place in the hydrophilic region of the molecule, the -ANN- sequence is not thought to be particularly activating, and so this degradation pathway would not have been predicted.

Crystallin-A (chicken) (173 residues)

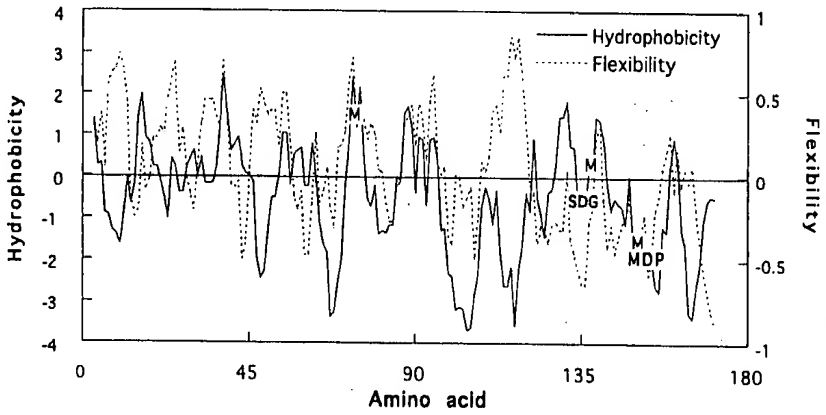
SEQUENCE

MDITIQHPWFKRALGPLIPSRLFDQFFGEGLLEYDLLPLFSSTISPYYRQSLFRSVLES-
GISEVRSDRDKFTIMLDVKHFSPEDLSVKIIDDFVEIHGKHSERQDDHGYISREFHRR-
YRLPANVDQSAITCSLSSDGMLTFSGPKVPSNMDPSHSERPIPVSREEKPTSAPSS

REACTIVE SITES

.N.(2)	.D.(14)	.M.(3)	.Q.(5)
123 ANV	2 MDI	74 IML	6 IQH
149 SNM	24 FDQ	138 GML	25 DQF
	35 YDL	150 NMD	50 RQS
	67 SDR		104 RQD
	69 RDK		126 DQS
	76 LDV		
	84 EDL		
	91 IDD		
	92 DDF		
	105 QDD		
	106 DDH		
	125 VDQ		
	136 SDG		
	151 MDP		

HYDROFLEX PLOT



PREDICTED REACTIVITY AND DEGRADATION OF α-CRYSTALLIN-A

This protein has only a few predicted hot spots and may be expected to show degradation at Asp-Gly or Met. This was not observed experimentally, however. Extraction and purification of α-crystallin-A from chicken eye lenses afforded a protein that showed microheterogeneity at position 149 (Voorter *et al.*, 1987) due to deamidation of Asn-149 within the -SNM- motif. This sequence is not predicted to be particularly reactive, in that Asn is not followed by either Gly or Ser. Indeed, the authors pointed out that the deamidation at this site is age-related and that only partial microheterogeneity was observed in 10-year-old chickens, but not observed in young chickens. Crystallin in eye lenses is known to have a negligible turnover rate, indicative that it takes 10 years at physiological temperature for even partial deamidation to occur. That

deamidation did not occur at other Asn residues is not surprising, in that the only other Asn in α-crystallin-A is located within an even less reactive motif (-ANV-). This protein contains an Asp that is predicted to show some degradation (Asp-136 within the -SDG- motif), but it is unlikely that the authors would have seen this with their method of high-voltage paper electrophoresis. No controls were carried out to show that Asn-149 undergoes deamidation under formulation conditions (neutral pH at 5–25°C within 2 years).

Cytochrome *c* (140 residues)

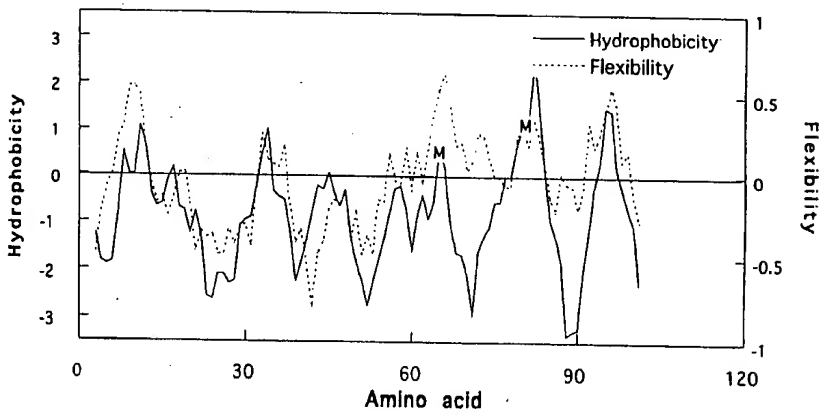
SEQUENCE

GDVEKGKKIFVQKCAQCHTVEKGGKHKTGPNLHGLFGRKTGQAPGFSYTDANKN-
KGITWGEETLMEYLENPKKYIPGTKMIFAGIKKKGEREDLIAYLKKATNE

REACTIVE SITES

.N.(5)	.D.(3)	.M.(2)	.Q.(3)
31 PNL	2 GDV	65 LME	12 VQK
52 ANK	50 TDA	80 KMI	16 AQC
54 KNK	93 EDL		42 GQA
70 ENP			
103 TNE			

HYDROFLEX PLOT



PREDICTED REACTIVITY AND DEGRADATION OF CYTOCHROME *c* (Cy I)

Inspection of the primary amino acid sequence for cytochrome *c* shows that it is devoid of traditional hot spots (Asn-Gly, Asn-Ser, Asp-Gly, and Gln-Gly), and so, *a priori*, this protein might be expected to be fairly stable, at least at 2–8°C at neutral pH. One of the earliest studies to detail protein microheterogeneity was reported by Flatmark on the reaction of cytochrome *c* in aqueous buffers (Flatmark, 1966). This protein showed microheterogeneity at Asn-103 near the C-termini (-TNE) after tryptic mapping. This region of cytochrome *c* is predicted to be both hydrophilic and flexible, but this Asn is located within a motif not thought to be reactive. Although this early report has some errors in the interpretation of the kinetic data (for example, neglecting pyro-Glu formation in the comparative analysis of Glu and Asn free amino acid reactivity, or the "visual" determination that the rate constant for reaction of Cy I is significantly less than for Cy II for the sequential reaction Cy I → Cy II → Cy III), it was found that the major site of microheterogeneity in Cy I was Asn-103 in the -TNE motif at the C-terminus. (It appears that the Cy II subfraction showing microheterogeneity was obtained by preparative work-up of tissue rather than as the degradation product of Cy I, and so the reaction of Asn-103 is considered a "work-up" deamidation reaction.) Data were reported at both 4 and 37°C and at several pHs ranging from 3 to 11. The pH rate profiles for reaction of Cy I suggest that several reactions may be occurring, in that the slope of these plots in the region of base catalysis has a slope significantly less than unity (when plotted as log *k* versus pH). Refitting the data to a standard log *k*–pH rate profile suggests that cytochrome *c* should exhibit a 2-year shelf life below pH ~7.5. Indeed, use of the kinetic data provided in this chapter to construct a typical log *k*–pH rate profile suggests that cytochrome *c* should exhibit a shelf-life in aqueous solution of 20 years or more at pH 6 (Fig. 2). Based on this, the -TNE motif is fairly unreactive, despite its ease of deamidation at high pH and 37°C.

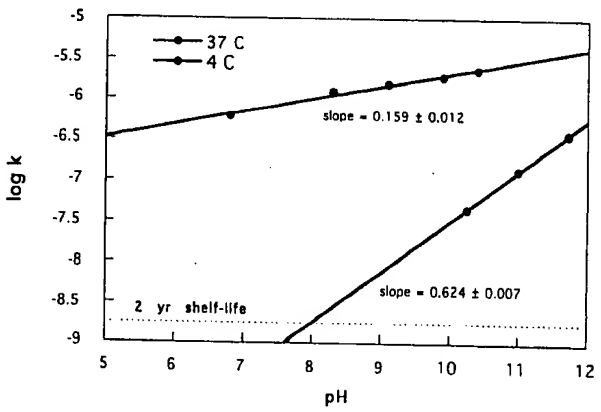


Figure 2. Log *k*–pH rate profiles for the reaction of Cy I to Cy II in aqueous buffers at 4° and 37°C. The slopes of values less than unity suggest that degradation may be occurring by several pH-dependent pathways. Linear extrapolation at 4°C suggests that the shelf life (due to deamidation) of cytochrome *c* will be >20 years at pH 6.

DNase (human) (260 residues)

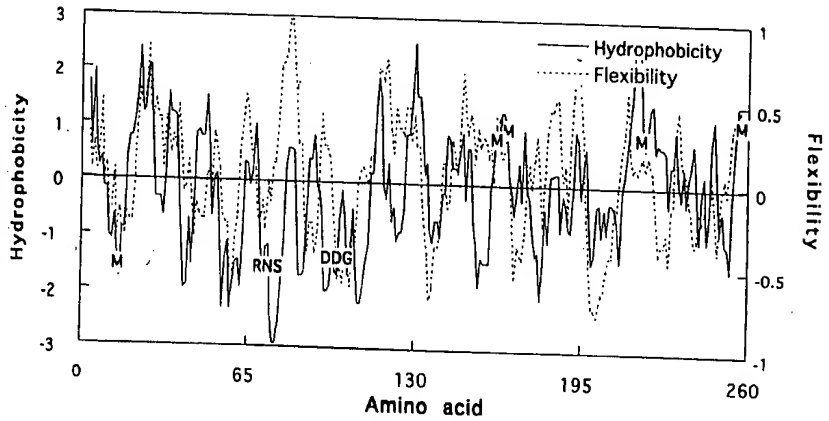
SEQUENCE

LKIAAFNIQTFGETKMSNATLVSYIVQILSRYDIALVQEVRDSHLTAVGKLLDNLNQ-
DAPDTYHYVVSEPLGRNSYKERYLFVYRPDQVSAVDSYYYDDGCEPCGNDTFNRE-
PAIVRFFSRFTEVREFAIVPLHAAPGDRVAEIDALYDVYLDVQEKWGLEDVMLMGD-
FNAGCSYVRPSQWSSIRLWTSPTFQWLIPDSADTTATPTHCAYDRIVVAGMLLRGA-
VVPDSALPFNFQAAYGLSDQLAQAISDHYPVEVMLK

REACTIVE SITES

.N.(9)	.D.(22)		.M.(5)	.Q.(11)
7 FNI	33 YDI	145 IDA	16 KMS	9 IQT
18 SNA	42 RDS	149 YDV	164 VML	27 VQI
54 DNL	53 LDN	153 LDV	166 LMG	38 VQE
56 LNQ	58 QDA	162 EDV	219 GML	57 NQD
74 RNS	61 PDT	168 GDF	258 VML	88 DQV
106 GND	87 PDQ	198 PDS		155 VQE
110 FNR	93 VDS	201 ADT		180 SQW
170 FNA	98 YDD	212 YDR		193 FQW
234 FNF	99 DDG	228 PDS		236 FQA
	107 NDT	243 SDQ		244 DQL
	139 GDR	251 SDH		247 AQA

HYDROFLEX PLOT



PREDICTED REACTIVITY AND DEGRADATION OF DNase

Deamidation in DNase occurs at the -RNS- motif, where it is expected that this Asn is the likely hot spot due to the presence of Ser on the C-terminal side, as well as the flanking polar

Arg. The hydropathy plot also supports this as the most likely site of deamidation, in that this motif is predicted to exist in a hydrophilic region of intermediate flexibility. There is another motif (-DDG- at Asp-99) that is also predicted to be a hot spot, in that it exists in a hydrophilic flexible region. Reaction at this site, however, has not been observed. The major degradation pathway of DNase at pH 5–8 in aqueous solution was found to be deamidation at Asn-74, giving the Asp and the iso-Asp variants. Modification at this site does not lead to complete inactivity, wherein the deamidated product exhibited ~50% of the original activity (Frenz, 1991; Cipolla *et al.*, 1994). Reaction at this site did not compromise a 2-year shelf life when the product was stored at 2–8°C.

Epidermal Growth Factor (EGF 1–48) (human) (48 residues)

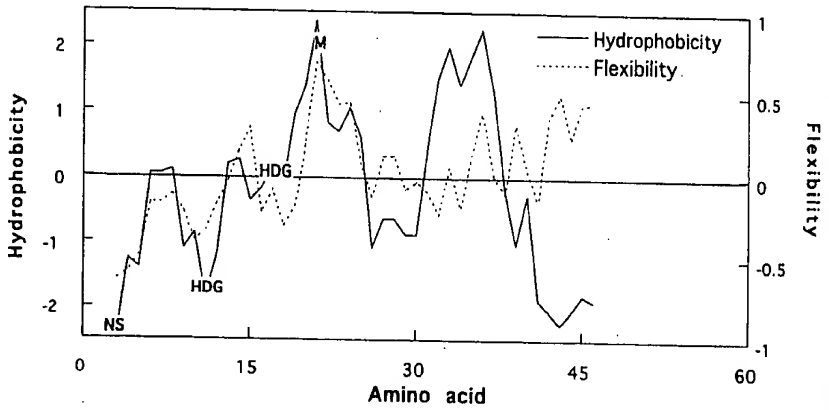
SEQUENCE

NSDSECPLSHDGYCLHDGVCMYIEALDKYACNCVVGYIGERCQYRDLK

REACTIVE SITES

.N.(1)	.D.(5)		.M.(1)	.Q.(1)
1 NS	3 SDS	27 LDK	21 CMY	43 CQY
32 CNC	11 HDG	46 RDL		
	17 HDG			

HYDROFLEX PLOT



PREDICTED REACTIVITY AND DEGRADATION OF EGF (HUMAN)

The hydroflex plot for EGF shows that there are two reactive hydrolytic sites (Asn-1 in the NS- motif and Asp-11 in the -HDG- motif). Oxidation of Met may also be a likely pathway for

degradation. An elegant study on the degradation of EGF showed that succinimide formation at Asp-11 was most prominent below pH 6, whereas deamidation of Asn-1 was the primary degradation pathway above pH 6 (Senderoff *et al.*, 1994). The relative contribution of oxidation was found to increase as the temperature was lowered. This study also included the pH rate profiles for both reactant loss and degradation product formation at both 4 and 30°C. Deamidation as the primary degradation pathway for EGF at neutral pH was recently confirmed in another study showing the effects of buffer ions and surfactants at higher temperatures (Son and Kwon, 1995).

Epidermal Growth Factor (murine) (53 residues)

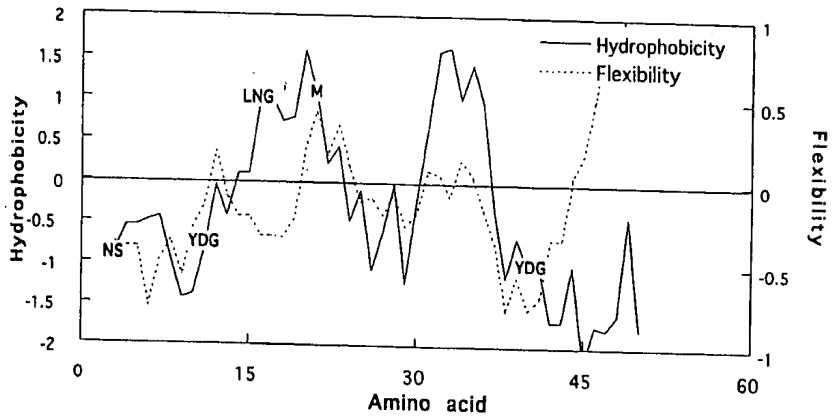
SEQUENCE

NSYPGCPSSYDGYCLNGGVCMHIESLDSYTCNCVIGYSGDGCQTRDLRWWQLR

REACTIVE SITES

.N.(2)	.D.(4)	.M.(1)	.Q.(2)
1 NS	11 YDG	21 CMH	43 CQT
16 LNG	27 LDS		51 WQL
32 CNC	40 GDG		
	46 RDL		

HYDROFLEX PLOT



PREDICTED REACTIVITY AND DEGRADATION OF EGF (MURINE)

The hydroflex plot for EGF shows that there are two reactive Asn (Asn-1 in the NS- motif and Asn-16 in the -LNG- motif). Insufficient data exist to predict which of these should react

fastest, because of the lack of reactivity data for N-terminal Asn residues. Reaction of murine EGF at pH 9 and 37°C for 48 hr afforded primarily reaction at Asn-1 (NS-), with a small amount of reaction at Asn-16 (-LNG-) (Galletti *et al.*, 1989). Unfolding the protein increased the amount of reaction at Asn-16. A similar reaction was observed at pH 7.4 at 22°C, where the observed half-life was approximately 500 hr (DiAugustine *et al.*, 1987); it was also reported that the half-life at pH ~13 and 22°C was 63 hr, suggesting that the rate of deamidation is not linearly proportional with hydroxide ion concentration (Tyler-Cross and Schirch, 1991). Insufficient experiments were carried out to determine the rate of reaction at pH 4.5–7.5 at 2–8°C.

Erythrocyte Protein 4.1 (human) (588 residues)

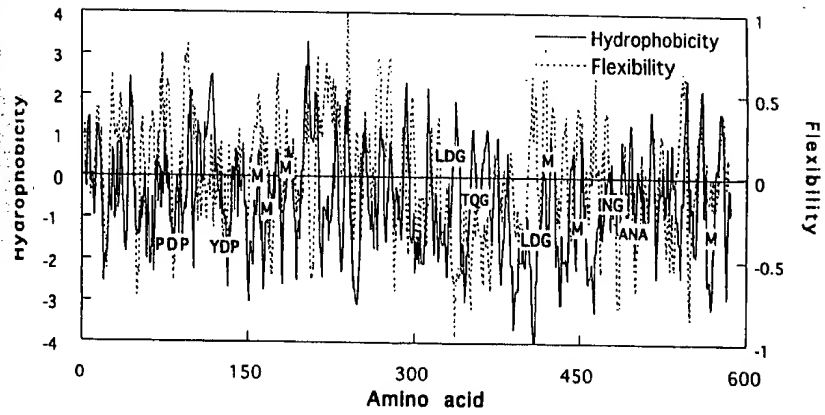
SEQUENCE

MHCKVSLLDDTVYECVVEKHAKGQDLLKRVCEHLNLLEEDYFGLAIWDNATSKT-
WLDSAKEIKKQVRGVPWNFTFNVKFYPPDPAQLTEDITRYYLCLQLRQDIVAGRLPC-
SFATLALLGSYTIQSELGDYDPELHGVDYVSDFKLAPNQTKELEEKVMELHKSYRS-
MTPAQADLEFLENAKKLSMYGVDLHKAKDLEGVDIILGVCSSGLLVYKDKLRINR-
FPWPKVLKISYKRSSFFIKIRPGEQEQYESTIGFKLPSYRAAKKLWKVCVEHHTFFRL-
TSTDTIPKSKFLALGSKFRYSGRTQAQTRQASALIDRPAPHFERTASKRASRSLDGAA-
AVDSADRSPRPTSAPAITQGQVAEGGVLDASAKKTVVPKAQKETVKAEVKKEDEPP-
EQAEPEPTEAWKKKRERLDGENIYIRHSNLMLEDLDKSQEEIKKHHASISELKKNF-
MESVPEPRPSEWDKRLSTHSPFRTLNINGQIPTGEGPPLVKTQTVTISDNANAVKSEI-
PTKDVPIVHTETKTITYEAAQTVKGGISETRIEKRIVITGDADIDHDQVLVQAIKEAK-
EQHPDMSVTKVVVHQETEIADE

REACTIVE SITES

.N.(14)	.D.(37)		.M.(6)	.Q.(25)	
35 LNL	9 LDD	316 IDR	159 VME	24 GQD	493 TQT
49 DNA	10 DDT	335 LDG	168 SMT	65 KQV	529 AQT
72 WNF	25 QDL	341 VDS	186 SMY	86 AQL	556 DQV
76 FNV	40 EDY	344 ADR	425 LML	99 LQL	560 VQA
149 PNQ	48 WDN	367 LDA	451 FME	102 RQD	568 EQH
180 ENA	57 LDS	391 EDE	572 DMS	125 IQS	581 HQE
221 INR	83 PDP	413 LDG		150 NQT	
416 ENI	90 EDI	428 EDL		172 AQA	
423 SNL	103 QDI	430 LDK		247 EQE	
449 KNF	130 GDY	463 WDK		249 EQY	
476 LNI	132 YDP	499 SDN		305 TQA	
478 ING	139 VDY	512 KDV		307 AQT	
500 DNA	143 SDF	549 GDA		310 RQA	
502 ANA	174 ADL	551 ADI		357 TQG	
	190 VDL	533 IDH		359 GQV	
	196 KDL	555 HDQ		379 AQK	
	201 VDI	571 PDM		396 EQA	
	216 KDK	587 ADE		433 SQE	
	284 TDT			480 GQI	

HYDROFLEX PLOT



PREDICTED REACTIVITY AND DEGRADATION OF ERYTHROCYTE PROTEIN 4.1

This protein has several hot spots of predicted reactivity, including an Asn-Gly motive. Isolation and purification of this large protein result in the selective deamidation at two sites, Asn-478 and Asn-502 (Inaba *et al.*, 1992). The first is unremarkable in that Asn-478 is adjacent to Gly (-ING-). Asn-502 is flanked by Ala on both sides (-ANA-), yielding an Asn that would be only weakly reactive based on model peptide studies. Indeed, the authors found that reaction of Asn-502 was much slower than at Asn-478, taking months for reaction to occur *in vivo*. No controls were carried out to show that reaction of Asn-502 occurred under formulation conditions in the absence of catalytic enzymes, nor was sufficient kinetic data presented (other than the reaction was slow) to permit an estimation of the reaction rate at 2–8°C.

Fibroblast Growth Factor, Acidic (human) (aFGF) (141 residues)

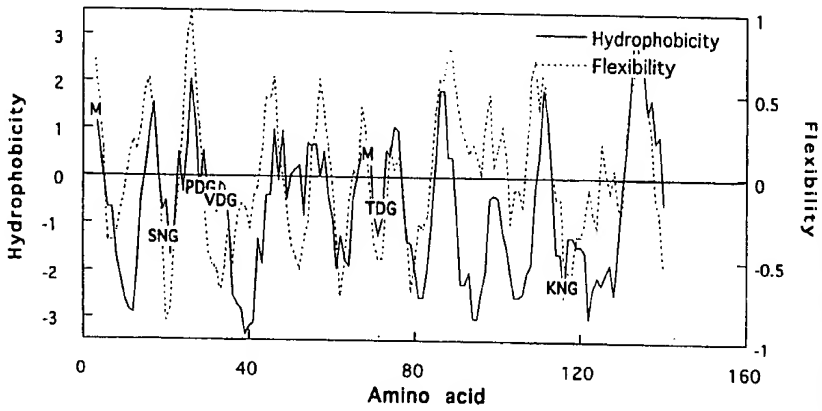
SEQUENCE

MFNLPPGNYKKPKLLYCSNGGHFLRILPDGTVDGTRDRSDQHIQLQLSAESVGEV-
YIKSTETGQYLAMDTDGLLYGSQTPNEECLFLERLEENHYNTYISKKHAEKNWFV-
GLKKNGSCKRGPRTHYGQKAILFLPLPVSSD

REACTIVE SITES

.N.(8)	.D.(7)	.M.(2)	.Q.(6)
3 FNL	29 PDG	1 MF	41 DQH
8 GNY	33 VDG	68 AMD	44 IQL
19 SNG	37 RDR		46 LQL
81 PNE	40 SDQ		64 GQY
93 ENH	69 MDT		78 SQT
96 YNT	71 TDG		128 GQK
107 KNW	141 SD		
115 KNG			

HYDROFLEX PLOT



PREDICTED REACTIVITY AND DEGRADATION OF ACIDIC FGF

Acidic FGF contains two Asn-Gly motifs that are predicted to be reactive, as well as a number of Asp-Gly residues. This molecule is known to be fairly reactive in solution, and so elegant formulations have been designed utilizing its stabilizing complexation with heparin to extend shelf life (Volkin and Middaugh, 1996). Some degradation studies have been carried out, wherein it was found that deamidation was one of the major degradation pathways. N-terminal sequence analysis showed that Asn-8 (-GNY-) was deamidated, but that Asn-19 was not (-SNG-). This is somewhat unusual, in that the Asn-Gly sequence is usually much more reactive than the Asn-Tyr sequence. The authors pointed out that Asn-8 is in a hydrophilic flexible region, possibly enhancing its reactivity. Conversely, Asn-19 is located in the heparin binding region for acidic-FGF, and this may contribute to its lack of reactivity. No degradation was reported for the Asn-Gly site near the C-termini, although the methods used (sequence analysis) were not developed to look at this region of the molecule. No oxidation of Met was reported, however oxidation at Cys leads to inactivation of the protein.

Fibroblast Growth Factor, Basic (human) (bFGF) (154 residues)

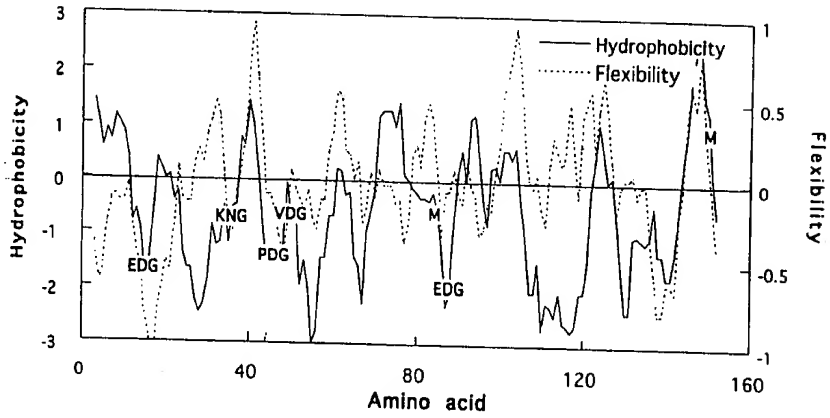
SEQUENCE

AAGSITTLPALPEDGGSGAFPPGHFKDPKRLYCKNGGFFLRIHPDGRVDGVREKSDP-
HIKLQLQAEERGVVSIKGVCANRYLAMKEDGRLLASKCVTDECFFFERLESNNYNT-
YRSRKYTSWYVALKRTGQYKLGSKTGPGQKAILFLPMSAKS

REACTIVE SITES

.N.(5)		.D.(7)		.M.(2)	.Q.(4)
36 KNG	113 YNT	15 EDG	57 SDP	85 AMK	63 LQL
80 ANR		28 KDP	88 EDG	151 PMS	65 LQA
110 SNN		46 PDG	99 TDE		132 GQY
111 NNY		50 VDG			143 GQK

HYDROFLEX PLOT



PREDICTED REACTIVITY AND DEGRADATION OF BASIC FGF

Degradation of bFGF occurs at Asp-15 within the -EDG- motif, where it is expected that this Asp is a likely hot spot due to the presence of Gly on the C-terminal side, as well as the flanking polar Glu. The hydropathy plot also supports this as the most likely site of succinimide formation, in that this motif is predicted to exist in a hydrophilic region of moderate flexibility. There also exists a -KNG- motif at Asn-36 that is likely to be a reactive hot spot, particularly at higher pH's. Asn-36 is found in a region of only intermediate hydrophobicity and flexibility, and so may be of reduced reactivity compared with a -KNG- motif found in smaller peptides. There is also another -EDG- motif found at Asp-88, although this Asp is predicted to be of lower reactivity because the regional flexibility is less than at Asp-15. When the stability of

bFGF was investigated at pH 6.5, the degradation product eluted sooner than the parent by HP-IEC, indicative of a more acidic deamidated product (although this was not confirmed with product analysis). The major degradation pathway of bFGF at pH 5 in aqueous solution was succinimide formation at Asp-15 (Shahrokh *et al.*, 1994). In addition, two truncated monomer forms were found as minor degradation products, due to cleavage at Asp-28-Pro and Asp-15-Gly. Modification at these sites did not lead to inactivity, where the cleaved or cyclized products remained bioactive in a heparin binding assay and in a cell proliferation assay. No evidence was found for oxidative degradation of Met within 13 weeks at pH 5 at 25°C. The stability of bFGF at 2–8°C was not addressed directly in this chapter, but mention was made that iso-Asp formation was less than 2% in 24 weeks at 4°C, which should be interpreted as a lower limit because the cyclic imide is also formed.

Glucagon (29 residues)

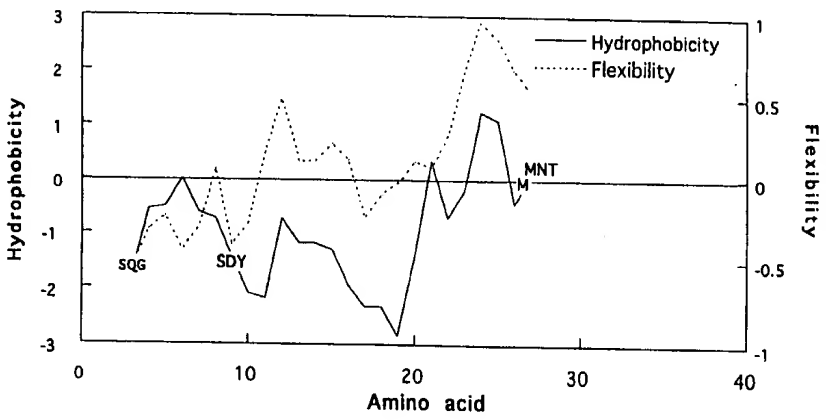
SEQUENCE

HSQGTFTSDYSKYLDSRRAQDFVQWLMNT

REACTIVE SITES

.N.(1)	.D.(3)	.M.(1)	.Q.(3)
28 MNT	9 SDY	27 LMN	3 SQG
	15 LDS		20 AQD
	21 QDF		24 VQW

HYDROFLEX PLOT



PREDICTED REACTIVITY AND DEGRADATION OF GLUCAGON

This peptide is not predicted to be very reactive, in that it is missing most of the traditional hot spots. Inspection of the hydroflex plot shows a Gln-Gly motive near the N-termini which may be expected to be mildly reactive. Methylation of glucagon to identify iso-Asp residues showed that glucagon contained some iso-Asp at Asp-9 and Asn-28 (Ota *et al.*, 1987). Both of these amino acids are located within motifs that are not expected to be reactive (-SDY- and -MNT-) based on data obtained in synthetic peptides. No control experiments were carried out to determine if the same degradation reaction occurs in pH 4.5–7.5 buffer. Of note, glucagon samples were boiled for a short time before carrying out the enzymatic maps, and the consequence of this preparative step on the degradation of glucagon was undetermined.

Granulocyte-Colony Stimulating Factor (G-CSF) (human) (175 residues)

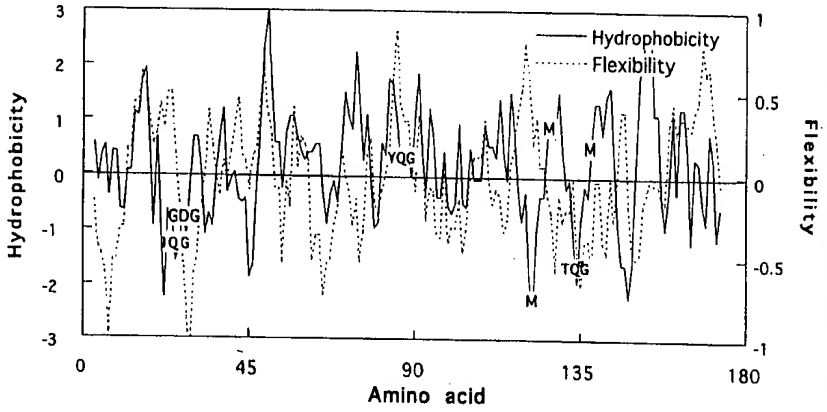
SEQUENCE

MTPLGPASSLPQSFLLKCLEQVRKIQGDGAALQEKLCATYKLCHPEELVLLGHSLGI-
PWAPLSSCPSQALQLAGCLSQLHSGLFLYQGLLQALEGISPELGPTLDTLQLDVADFA-
TTIWQQMEELGMAPALQPTQGAMPAFASAFQRRAGGVLVASHLQSFLEVSYRVLRH-
LAQP

REACTIVE SITES

.N.(0)	.D.(4)	.M.(3)	.Q.(17)
	28 GDG	122 QME	12 PQS
	105 LDT	127 GMA	21 EQV
	110 LDV	138 AMP	26 IQG
	113 ADF		33 LQE
			68 SQA
			71 LQL
			78 SQL
			87 YQG
			91 LQA
			108 LQL
			120 WQQ
			121 QQM
			132 LQP
			135 TQG
			146 FQR
			159 LQS
			174 AQP

HYDROFLEX PLOT



PREDICTED REACTIVITY AND DEGRADATION OF GCSF

This molecule is devoid of Asn and has only a few predicted hot spots such as Asp-Gly (cyclization and iso-Asp formation), Gln-Gly (deamidation), or Met (oxidation). The degradation pathways of GCSF in aqueous solution have been determined, wherein it was found that the predominant site of deamidation was at Gln-21 (in the -EQV- motif), and oxidation at Met-122 and at either Met-127 or Met-138 (these residues are in the same peptide in the tryptic digest and so differentiation has not been made) (Herman *et al.*, 1995). Even though Gln-21 is located in a region of predicted hydrophilicity, deamidation at Gln-21 is unexpected because the -EQV- motif is not a traditional hot spot based on the deamidation of Gln in small peptides.

- **Growth Hormone (bovine) (191 residues)**

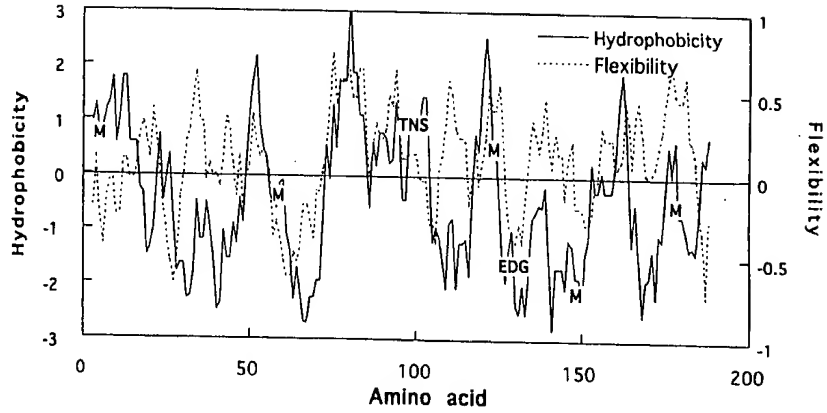
SEQUENCE

AFPAMSLSGLFANAVLRAQHLHQLAADTSKEFERTYIPEGQRYSIQNTQVAFCFSET-
MPAPTGKNEAQQKSDLELLRISLLLIQSWLGPLQFLSRVFTNSLVFGTSDRVYEKLK-
DLEEGILALMRELEDGTPRRGQILKQTYDKFDTNMRSDDALLKNYGLLSCFRKDL-
HKTETYLRVMKCRRFGEASCAF

REACTIVE SITES

.N.(6)	.D.(10)	.M.(5)	.Q.(11)
13 ANA	27 ADT	143 YDK	5 AMS	19 AQH	69 QQK
47 QNT	72 SDL	146 FDT	58 TMP	23 HQL	84 IQS
65 KNE	107 SDR	152 SDD	124 LMR	41 GQR	91 LQF
99 TNS	115 KDL	153 DDA	149 NMR	46 IQN	136 GQI
148 TNM	129 EDG	168 KDL	179 VMK	49 TQV	140 KQT
158 KNY				68 AQQ	

HYDROFLEX PLOT



PREDICTED REACTIVITY AND DEGRADATION OF BOVINE GROWTH HORMONE

The primary sequence of bovine growth hormone has two potential hot spots for hydrolytic degradation, where Asn-99 in the -TNS- motif is the most likely. The other hot spot is Asp-129, although it is likely that reactivity at this site at pH 7.4 will be much slower than at Asn-99. [It has been reported that Asp-129 is the predominant site of reaction in porcine somatotropin under acidic conditions giving the cyclic imide (Violand *et al.*, 1992).] Although the Asn-99 site is a predicted hot spot based on primary sequence alone, it is not predicted to be a hot spot based on the hydroflex plot, in that Asn-99 lies in a hydrophobic region of little flexibility. In contrast, Asp-129 lies in a hydrophilic flexible region. It has been shown that the primary site of degradation in bovine (and porcine) growth hormone at pH 7.4 was deamidation at Asn-99 to give predominantly iso-Asp-99 (Violand *et al.*, 1990). The authors pointed out that it is likely that reaction may occur at other sites in bovine growth hormone, but may not be resolvable under their HPLC conditions. These studies were carried out at 37°C, so an estimate of the reaction rate at 2–8°C could not be made from these studies. There is sufficient data on bovine growth hormone showing conclusively that bovine growth hormone undergoes different degradation pathways under "work-up" conditions and under "formulation" conditions,

even at the same pH and temperature. For example, the major reaction site in purified pituitary bovine growth hormone in aqueous buffer was Asn-99, but the predominant variants isolated from work-up samples were Asn-13 and Asn-148 in one study (Violand *et al.*, 1990) and Asp-129 in another (Wood *et al.*, 1989). Although reactivity at Asp-129 is not unexpected (see above), neither the Asn-13 (-ANA-) nor Asn-148 (-TNM-) is predicted to reactive based on primary sequence and hydroflex plot analysis. This is another example showing that deamidation or iso-Asp formation under work-up conditions should not be used to predict the site(s) of major degradation in typical pH 5–7 aqueous formulations. The interspecies variation in GH primary amino acid sequence is given in Scheme 10 for reference.

```
GH-b     1 AFPAMSLSGLFANAVLRAQHLHQLAADTSKEFERTYIPEGQRYS-IQNTQ
GH-h     1 -FPTIPLSRLFDNAMLRAHRLHQLAFDTYQEFEEAYIPKEQKYSFLQNPQ
GH-p     1 -FPAMPLSSLFANAVLRAQHLHQLAADTYKEFERTYIPEGQRYS-IQNAQ

GH-b    50 VAFCFSETMPAPTGKNEAQQKSDLELLRISLLLIQSWLGPLQFLSRVFTN
GH-h    50 TSLCFSESIPTPSNREETQQKSNLELLRISLLLIQSWLEPVQFLRSVFAN
GH-p    49 AAFCFSETIPAPTGKDEAQQRSDVELLRISLLLIQSWLGPVQFLSRVFTN

GH-b   100 SLVFGTSD-RVYEKLKDLEEGILALMRELEDGTPRRGQILKQTYDKFDTN
GH-h   100 SLVYGASDSNVYDLLKDLEEGIQTLMGRLEDGSPRTGQIFKQTYSKFDTN
GH-p    99 SLVFGTSD-RVYEKLKDLEEGIQALMRELEDGSPRAGQILKQTYDKFDTN

GH-b   149 MRSDDALLKNYGLLSCFRKDLHKTETYLRVMKCRRFGEASCAF
GH-h   150 SHNDDALLKNYGLLYCFRKDMDKVETFLRIVQCRS-VEGSCGF
GH-p   148 LRSDDALLKNYGLLSCFKKDLHKAETYLRVMKCRRFVESSCAF
```

Scheme 10. Primary amino acid sequences of bovine, human, and porcine growth hormones.

• Growth Hormone (human) (191 residues)

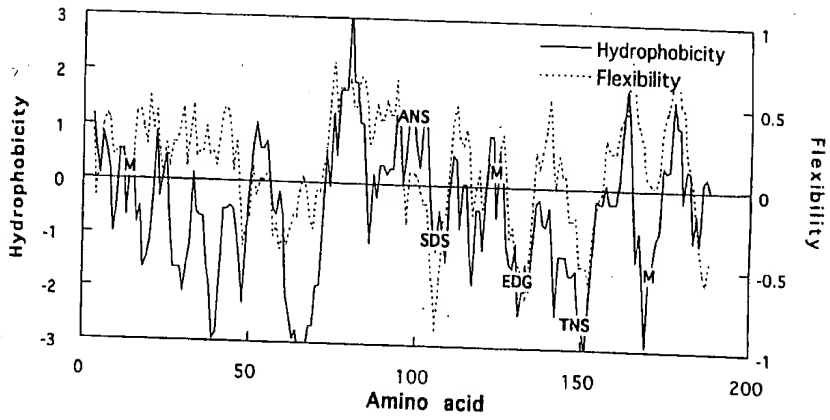
SEQUENCE

FPTIPLSRLFDNAMLRAHRLHQLAFDTYQEFEEAYIPKEQKYSFLQNPQTSLCFSESIP-
TPSNREETQQKSNLELLRISLLLIQSWLEPVQFLRSVFANSLVYGASDSNVYDLLKDL-
EEGIQTLMGRLEDGSPRTGQIFKQTYSKFDTNSHNDDALLKNYGLLYCFRKDMDKV-
ETFLRIVQCRSVEGSCGF

REACTIVE SITES

.N.(9)		.D.(11)		.M.(3)	.Q.(13)	
12 DNA	152 HND	11 FDN	153 NDD	14 AML	22 HQL	84 IQS
47 QNP	159 KNY	26 FDT	154 DDA	125 LMG	29 YQE	91 VQF
63 SNR		107 SDS	169 KDM	170 DMD	40 EQK	122 IQT
72 SNL		112 YDL	171 MDK		46 LQN	137 GQI
99 ANS		116 KDL			49 PQT	141 KQT
109 SNV		130 EDG			68 TQQ	181 VQC
149 TNS		147 FDT			69 QQK	

HYDROFLEX PLOT



PREDICTED REACTIVITY AND DEGRADATION OF HUMAN GROWTH HORMONE

There are three likely hot spots for hydrolytic degradation in hGH: Asn-99, Asp-130, and Asn 149. Degradation of hGH occurred primarily at Asn-149 and Asp-130, as might be expected in that Asn is next to Ser and Asp is next to Gly. The hydropathy plot also supports this as the most likely site of degradation, in that these motifs exist in a hydrophilic region of good flexibility. Peptide chain flexibility is probably quite important for the deamidation of Asn-149 in human growth hormone (Johnson *et al.*, 1989b). The structure of human growth hormone is likely to be similar to porcine growth hormone, which is poorly ordered in the region of residues 128 to 151 (Abdel-Meguid *et al.*, 1987). Asn-99 has a similar motif (-ANS-) as Asn-149 (-TNS-), and yet Asn-99 does not undergo reaction. An elegant explanation for this has been given by comparing the bovine and human sequences of growth hormone and then rationalizing the decreased reactivity at this site by an unfavorable conformational structure near Asn-99. Often this in-depth explanation is not possible because the 3D structure is unavailable, so it would be useful if this lack of reactivity could be predicted based on primary sequence hydrophobicity calculations. Indeed, inspection of the hydropathy and flexibility plots suggests that Asn-149 should be reactive (in that it exists in a hydrophilic region of good flexibility), whereas the Asn-99 exists in a hydrophobic region of lower flexibility and thus may be removed from the solvent and less available for reaction. The major degradation pathway of hGH at pH 6 in aqueous solution was found to be deamidation at Asn-149, with minor degradation pathways including cyclic imide and iso-Asp formation at Asp-130 (Teshima *et al.*, 1991b), and oxidation at Met-14 and Met-125 (Teshima and Canova-Davis, 1992). None of these reactions, nor their sum, compromised the shelf life of the liquid growth hormone formulation, having a shelf life of at least 18 months at 2–8°C. This formulation contained a preservative, as well as Tween 20. In another study, the degradation products of hGH were also determined after incubation at pH 7.4 and 37°C, giving largely deamidation at Asn-149 to form the iso-Asp and Asp degradation products. A small amount of deamidation was found at Asn-152. Iso-Asp formation at Asp-130 was also observed, but not deamidation at Asn-99, a

similar Asn sequence of -ANS-. It has also been reported that hGH forms the N-terminal diketopiperazine product during fermentation and/or work-up, although this is not a degradation product in the final formulation (Battersby *et al.*, 1994).

Growth Hormone (porcine) (190 residues)

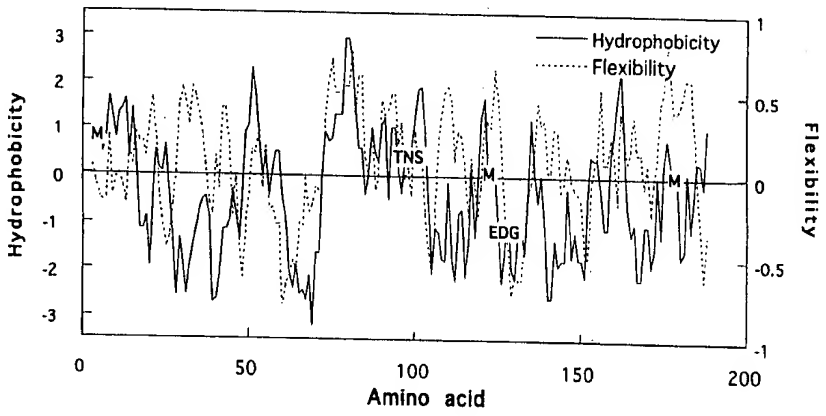
SEQUENCE

FPAMPLSSLFANAVLRAQHLHQLAADTYKEFERTYIPEGQRYSIQNAQAAFCFSETIP-
APTGKDEAQQRSDVELLRISLLLIQSWLGPVQFLSRVFTNSLVFGTSDRVYEKLKDL-
EEGIQALMRELEDGSPRAGQILKQTYDKFDTNLRSDDALLKNYGLLSCFKKDLHKA-
ETYLRVMKCRRFVESSCAF

REACTIVE SITES

.N.(5)	.D.(11)		.M.(3)	.Q.(12)	
12 ANA	26 ADT	142 YDK	4 AMP	18 AQH	68 QQR
46 QNA	64 KDE	145 FDT	123 LMR	22 HQL	83 IQS
98 TNS	71 SDV	151 SDD	178 VMK	40 GQR	90 VQF
147 TNL	106 SDR	152 DDA		45 IQN	120 IQA
157 KNY	114 KDL	167 KDL		48 AQA	135 GQI
	128 EDG			67 AQQ	139 KQT

HYDROFLEX PLOT



PREDICTED REACTIVITY AND DEGRADATION OF PORCINE GROWTH HORMONE

In contrast to human GH, there are only a few predicted hot spots for hydrolytic degradation in porcine GH, at Asn-98, Asp-128, and the Met residues. Degradation of pGH is predicted to occur primarily at Asn-98, as might be expected in that Asn is next to Ser (Violand *et al.*, 1990). This residue is in a moderately hydrophobic region, and so degradation might be expected to be slower than if it were in a hydrophilic region. The major degradation pathway of pGH in aqueous solution was found to be deamidation at Asn-98, with other degradation pathways at residues Cys-180–Cys-188 and Cys-52–Cys-163 (McCrossin *et al.*, 1994). This study was carried out at pH 9 to effect faster reaction rates, and this higher pH may be the reason that reaction occurred at the Cys-Cys bonds. Under these conditions, reaction at Asn-98 gave iso-Asp-98 and Asp-98 in a 3:1 ratio.

Growth Hormone Releasing Factor (GHRF) Variant (human) (32 residues)

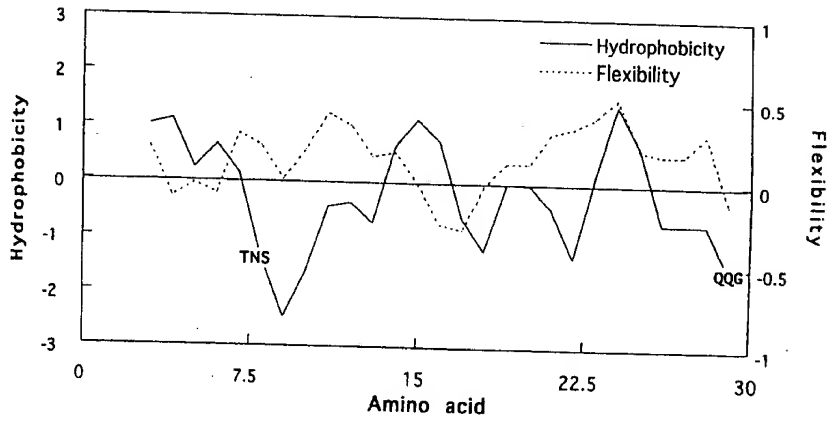
SEQUENCE

YADAIFTNSYRKVLGQLSARKLLQDILSRQQG

REACTIVE SITES

.N.(1)	.D.(2)	.M.(0)	.Q.(4)	
8 TNS	3 ADA		16 GQL	30 RQQ
	25 QDI		24 LQD	31 QQG

HYDROFLEX PLOT



PREDICTED REACTIVITY AND DEGRADATION OF (Leu-27) GHRF (1–32) NH_2

The hydroflex plot for this GHRF variant shows that Asn-8 resides in a hydrophilic region or intermediate flexibility (although peptides such as GHRF may show flexibility throughout because of their small size), and thus Asn-8 may be expected to be a reactive site. Of secondary predicted reactivity is the C-terminal Glu, in that Glu-Gly typically reacts somewhat slower than Asn-Ser. Reaction of GHRF in aqueous solution at pH 7.4 and 37°C gave primarily reaction at Asn-8 (-TNS-) (Friedman *et al.*, 1991). Studies have been carried out using modified bovine GHRF analogues (for example, substitution of Gly-15 with Pro-15 or Ala-15 to disrupt the helical structure in the helical region near Asn-8), and these showed altered rates of deamidation (Stevenson *et al.*, 1993). Insufficient experiments were carried out to determine the rate of reaction at pH 4.5–7.5 or at 2–8°C. The parent molecule has Met at position 27, and has been nonenzymatically oxidized to give Met sulfoxide (Campbell *et al.*, 1990).

Hemoglobin (human) (146 residues)

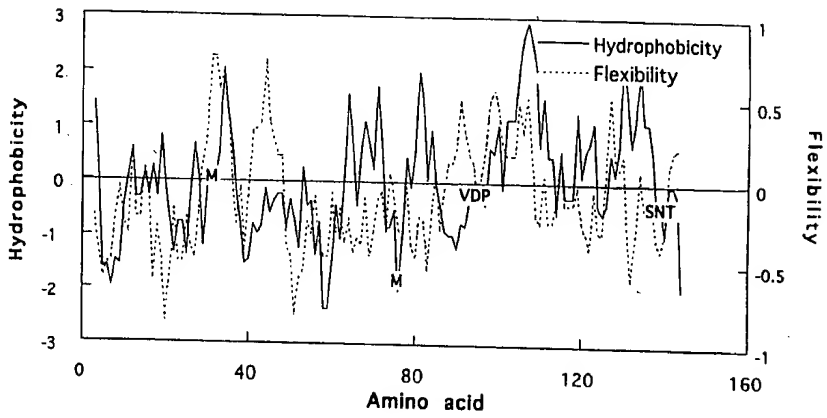
SEQUENCE

VLSPADKTNVKAAWGKVGAHAGEYGAEALERMFLSFPTTKTYFPHFDLSHGSAQ-
VKGHGKKVADALTNAVAHVDDMPNALSALSDLHAHKLRVDPVNFKLLSHCLLVTL-
AAHLPAEFTPAVHASLDKFLASVSTVLTSNTVKLQPR

REACTIVE SITES

.N.(5)	.D.(8)	.M.(2)	.Q.(2)
9 TNV	6 ADK	32 RMF	54 AQV
68 TNA	47 FDL	76 DMP	144 LQP
78 PNA	64 ADA		
97 VNF	74 VDD		
139 SNT	75 DDM		
	85 SDL		
	94 VDP		
	126 LDK		

HYDROFLEX PLOT



PREDICTED REACTIVITY AND DEGRADATION OF HEMOGLOBIN (WAYNE)

This variant of hemoglobin has few predicted reactive sites, the most likely being cleavage of Asp-Pro (-VDP-) at low pH, or oxidation of Met. Isolation and purification of this alpha-chain variant gives two forms, where the microheterogeneity was found at Asn-139 within the internal sequence -SNT- (Seid-Akhavan *et al.*, 1976). This motif is considered fairly unreactive, based on data obtained in small peptides (Tyler-Cross and Schirch, 1991). No controls were carried out to show that reaction of Asn-139 occurs under formulation conditions (for example, in the absence of catalytic enzymes), nor was sufficient kinetic data presented (other than the reaction was slow) to permit an estimation of the reaction rate at 2–8°C. This is another example of deamidation occurring at site other than Asn-Gly or Asn-Ser, but no evidence showing that the reaction proceeds rapidly by a nonenzymatic reaction.

Hirudin (65 residues)

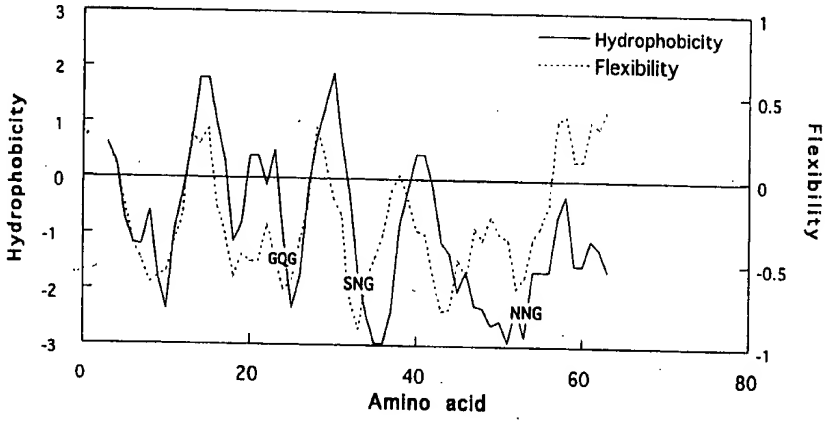
SEQUENCE

VVYTDCTESGQNLCLCEGSNVCGQGNKCILGSNGEKNQCVTGEGTPKPQSHNNG-
DFEEIP

REACTIVE SITES

.N.(5)	.D.(4)	.M.(0)	.Q.(5)
12 QNL	5 TDC		11 GQN
20 SNV	55 GDF		24 GQG
26 GNK			38 NQC
33 SNG			49 PQS
37 KNQ			65 LQ
52 HND			
53 NNG			

HYDROFLEX PLOT



PREDICTED REACTIVITY AND DEGRADATION OF HIRUDIN

Hirudin is a glycoprotein under development as an antithrombotic agent. This small protein contains two Asn-Gly residues in flexible, hydrophilic regions, and may be expected to undergo some cyclization and iso-Asp formation. Mass spectral degradation product studies on hirudin showed that Asn-33 underwent cyclic imide formation to form the Q4 variant with resultant Asp-33 formation, as well as a Q5 variant with Asp-53 (Grossenbacher *et al.*, 1993). These studies were carried out under slightly acidic conditions, which may promote degradation. Further, these product studies were carried out under harsh conditions (180-fold molar excess DTT, followed by pyridylation at pH 8.3 and enzymatic degradation at pH 7.8 at 37°C), making it difficult to determine if these degradation pathways would compromise hirudin shelf-life in a neutral pH aqueous formulation at 2–8°C. Another study of the hirudin variant 2 (rHV-Lys-47) showed that Asn-33 and Asn-53 (again both found within the Asn-Gly motif) were altered, either in the fermentation process or during the work-up (no stability data under formulation conditions were reported); it was determined that the variants were stable succinimide intermediates (Bischoff *et al.*, 1993). Based on these studies, both Asn-33 and Asn-53 in hirudin are very susceptible to cyclic imide formation.

Histone (102 residues)

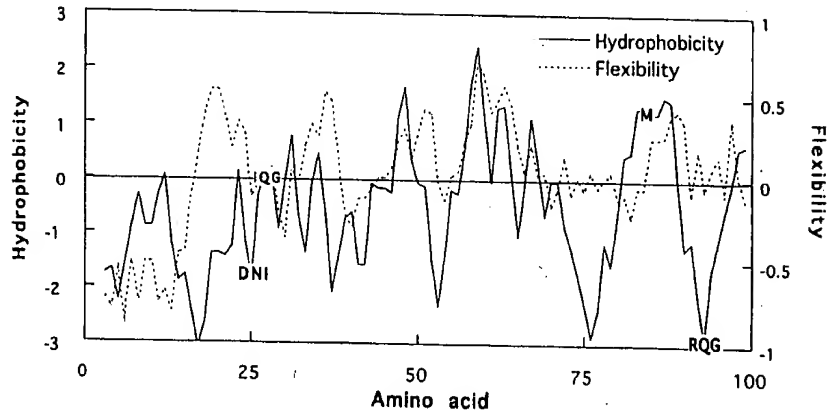
SEQUENCE

SGRGKGGKGLGKGGAKRHRKVLRDNIQGITKPAIRRLARRGGVKRISGLIYEETRG-
VLKVFLENVIRDAVTYTEHAKRKTVTAMDVVYALKRQGRTLYGFGG

REACTIVE SITES

.N.(2)	.D.(3)	.M.(1)	.Q.(2)
25 DNI	24 RDN	84 AMD	27 IQG
64 ENV	68 RDA		93 RQG
	85 MDV		

HYDROFLEX PLOT



PREDICTED REACTIVITY AND DEGRADATION OF HISTONE

The amino acid sequence of histone from several species was carried out as part of a study on histone evolution (Hayashi *et al.*, 1982). The authors found that histone-H4 (one of the fractionated histones in this study) showed microheterogeneity at Asn-25 after purification, where some Asp-25 was detected. Asn-25 resides within the -DNI- primary sequence motif, one that is thought to be fairly unreactive because of the neighboring Ile. Isolation of histone-H4 was done using several steps that might promote deamidation, including storage of the denatured protein with 2-mercaptoethanol at pH 8 and 40°C, an extraction and separation at pH 2.8, and purification using a pyridine–performic acid gradient at 55°C. No controls were carried out to determine if this unusual deamidation reaction occurred during the work-up or if the deamidation would occur in pH 4.5–7.5 formulation buffer.

Hypoxanthine-Guanine Phosphoribosyltransferase (HXGT) (217 residues)

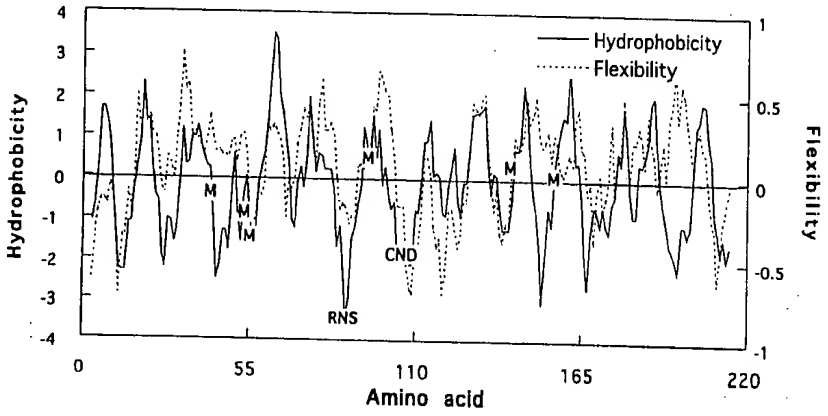
SEQUENCE

ATRSPGVVISDDEPGYDLDLFCIPNHYAEDLERVFIPHGLIMDRTERLARDVMKEMG-
GHHIVALCVLKGGYKFFADLLDYIKALNRNSDRSIPMTVDFIRLKSYCNDQSTGDIK-
VIGGDDLSTLTGKNVLIVEDIIDTGKTMQTLLSLVRQYNPKMVKVASLLVKRTPRSV-
GYKPDFVGFEIPDKFVVGYALDYNEYFRDLNHVCVISETGKAKYKA

REACTIVE SITES

.N.(8)	.D.(21)		.M.(6)	.Q.(3)
25 PNH	11 SDD	107 NDQ	42 IMD	108 DQS
85 LNR	12 DDE	112 GDI	53 VMK	143 MQT
87 RNS	17 YDL	119 GDD	56 EMG	151 RQY
106 CND	19 LDL	120 DDL	94 PMT	
128 KNV	30 EDL	134 EDI	142 TMQ	
153 YNP	43 MDR	137 IDT	156 KMV	
195 YNE	51 RDV	176 PDF		
202 LNH	76 ADL	184 PDK		
	79 LDY	193 LDY		
	89 SDR	200 RDL		
	97 VDF			

HYDROFLEX PLOT



PREDICTED REACTIVITY AND DEGRADATION OF HXGT

This protein has several Met residues, suggesting that this may be one of the predominant degradation pathways. In addition, it has an Asn-Ser motif that may be expected to be mildly reactive. Isolation and purification of HXGT from normal human erythrocytes afforded a tetrameric product that showed heterogeneity after tryptic digestion and peptide mapping (Wilson *et al.*, 1982). The heterogeneity was localized to a peptide spanning Ser-103 to Lys-114, which encompasses Asn-106 within the -CND- motif. Although the work-up used in this paper included strong acid (9% formic acid), sufficient control experiments were carried out to show that this deamidation was not due to the work-up but to *in vivo* deamidation. No control experiments were carried out to show that the same deamidation reaction occurs in pH 4.5–7.5 buffer. No oxidation was reported.

Insulin (human)

SEQUENCE (A CHAIN) (21 residues)

GIVEQCCTSICSLYQIENYCN

SEQUENCE (B CHAIN) (30 residues)

FVNQHLCGSHLVEALYLVCGERGFFYTPKT

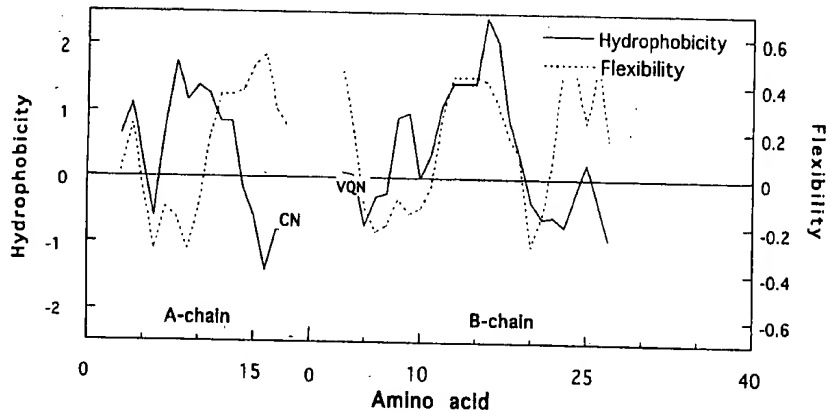
REACTIVE SITES (A CHAIN)

.N.(2)	.D.(0)	.M.(0)	.Q.(2)
18 ENY			5 EQC
21 CN			15 YQI

REACTIVE SITES (B CHAIN)

.N.(1)	.D.(0)	.M.(0)	.Q.(1)
3 VNQ			4 NQH

HYDROFLEX PLOT



PREDICTED REACTIVITY AND DEGRADATION OF HUMAN INSULIN

Insulin contains three Asn and two Gln that might be available for hydrolytic degradation. None of these residues, however, are predicted to be particularly susceptible to degradation based on their primary sequence, in that the motifs of highest reactivity, -XNG-, -XNS-, -XDG-, and -XQG-, are absent in insulin. Of the three Asn motifs in insulin, the -VNQ- would be predicted to be more reactive than the -ENY- or the -CN motifs, based on the deamidation rates in model peptides (Robinson and Rudd, 1974). Inspection of the hydropathy plots for insulin shows that the Asn-3 residue is in a region of intermediate hydrophobicity and flexibility. The major degradation pathway of insulin at neutral pH was deamidation of an Asn-3 in the B chain (Asn-B3), giving a mixture of Asp-3 and iso-Asp-3 (Brange *et al.*, 1992). The stability data also suggested that this deamidation at neutral pH is fairly slow, where only 0.05% per month was lost, corresponding to a shelf life of several years. Although of limited utility for the prediction of insulin stability at neutral pH, it was also noted that deamidation under acidic conditions occurred predominantly at Asn-21 in the A chain (Asn-A21) (Darrington and Anderson, 1994, 1995), where the reaction proceeded via rate-limiting formation of a cyclic anhydride intermediate.

• Insulin-like Growth Factor I (IGF-I) (70 residues)

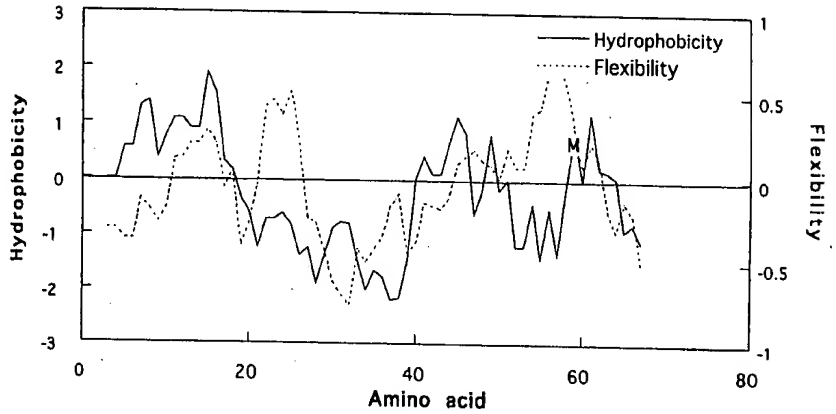
SEQUENCE

GPETLCGAELVDALQFVCGDRGFYFNKPTGYGSSSRRAPQTGIVDECCFRSCDLRR-LEMYCAPLKPAKSA

REACTIVE SITES

.N.(1)	.D.(4)	.M.(1)	.Q.(2)
26 FNK	12 VDA	59 EMY	15 LQF
	20 GDR		40 PQT
	45 VDE		
	53 CDL		

HYDROFLEX PLOT



PREDICTED REACTIVITY AND DEGRADATION OF IGF-I

Inspection of the primary sequence shows that IGF-I is missing the traditional hot spots for hydrolytic degradation, in that it is missing Asn-Gly, Asn-Ser, Asp-Gly, Asp-Pro, and Gln-Gly. There is a single Met (Met-59) found in a fairly hydrophobic region of low flexibility. Based on the primary amino acid sequence and the hydroflex plot, IGF-I is predicted to be a stable protein to hydrolytic and oxidative degradation. The major degradation route for IGF-I at pH 6 was found to be oxidation at Met-59; there was also some evidence for minor amounts of the des-Gly-Pro product formed by diketopiperazine formation (which is favored by Pro at position 2) (Poulter *et al.*, 1990). The sum of these degradation products did not compromise the shelf life of the product, where IGF-I was stable for more than 2 years at 2–8°C.

Insulinotropin (26 residues)

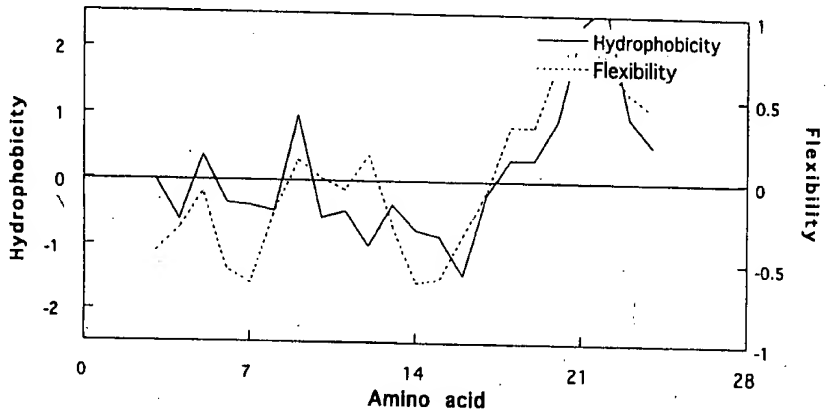
SEQUENCE

GTFTSDVSSYLEGQAAKEFIAWLVKG

REACTIVE SITES

.N.(0)	.D.(1)	.M.(0)	.Q.(1)
	6 SDV		14 GQA

HYDROFLEX PLOT



PREDICTED REACTIVITY AND DEGRADATION OF INSULINOTROPIN

Insulinotropin does not have the traditional hot spots and so is predicted to be fairly stable in aqueous solution. As an aside, this peptide has a highly hydrophobic C-terminal region and may be expected to show adsorption to surfaces and filters (Brophy and Lambert, 1994). This peptide is formulated in aqueous solution at pH containing 22.6% dextran to promote once-a-day subcutaneous injection. Under these conditions it was found that this peptide degraded fairly rapidly (t_{90} = ~40 hr at 25°C) giving biphasic kinetics (Heller and Qi, 1994). Excipient and degradation studies suggest that the Trp moiety is the reactive site in this peptide, corroborated by a significant loss in the absorption spectra at 300 nm.

Interferon-alpha-2b (human) (IFN-α-2b) (165 residues)

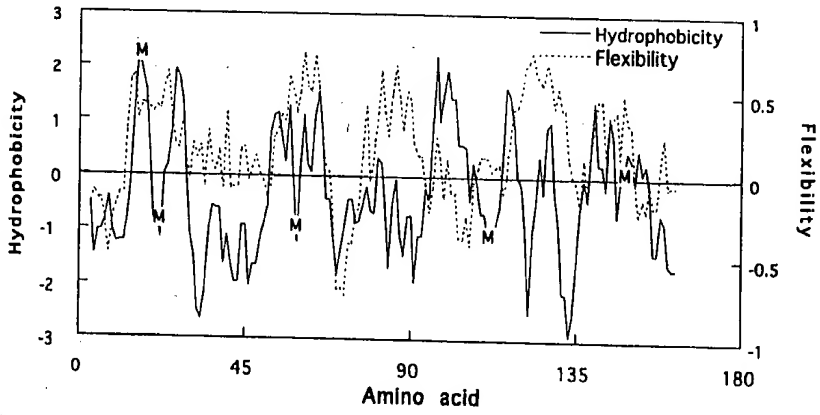
SEQUENCE

CDLPQTHSLGSRRTLMLLAQMRRISLFSCLKDRHDFGFPQEEFGNQFQKAETIPVLH-
EMIQQIFNLFSTKDSSAAWDETLLDKFYTELYQQLNDLEACVIQGVGVTETPLMKE-
DSILAVRKYFQRITLYLKEKKYSPCAWEVVRAEIMRSFSLSTNLQESLRSKE

REACTIVE SITES

.N.(4)	.D.(8)	.M.(5)	.Q.(12)
45 GNQ	2 CDL	16 LML	5 PQT
65 FNL	32 KDR	21 QMR	20 AQM
93 LND	35 HDF	59 EMI	40 PQE
156 TNL	71 KDS	111 LMK	46 NQF
	77 WDE	148 IMR	48 FQK
	82 LDK		61 IQQ
	94 NDL		62 QQI
	114 EDS		90 YQQ
			91 QQL
			101 IQG
			124 FQR
			158 LQE

HYDROFLEX PLOT



PREDICTED REACTIVITY AND DEGRADATION OF IFN-α-2b

This protein does not have any hydrolytic hot spots, so the main degradation routes (if any) would likely be due to Met oxidation. The Met-111 variant was isolated by RP-HPLC and

identified by tryptic mapping and mass spectral studies (Gitlin *et al.*, 1995). The oxidation of Met-111 did not affect the biological activity of this protein. Further, this variant was observed after fermentation and likely was not a formulation degradation product. Upon storage of IFN-α-2b in pH 7.2 phosphate buffer, there was some evidence for deamidation upon high-temperature thermal stress studies, but no extrapolation was made to determine the extent of deamidation at 2–8°C. These thermal stress studies did not increase the rate of Met oxidation at neutral pH.

Interferon-beta (IFN-β) (166 residues)

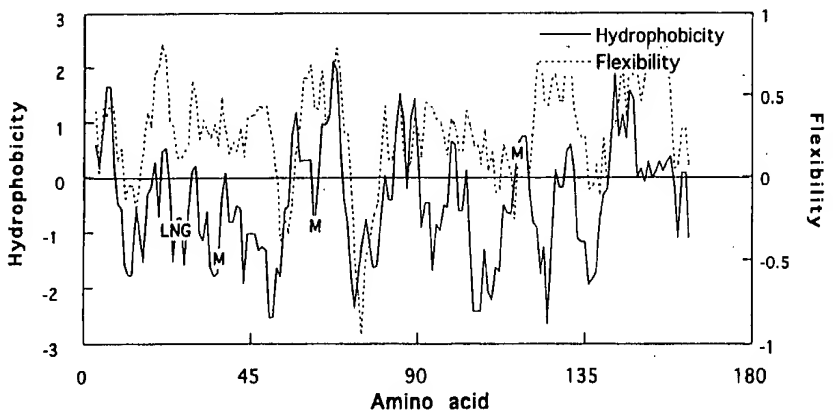
SEQUENCE

MSYNLLGFLQRSSNFQCQKLLWQLNGRLEYCLKDRMNFDIPEEIKQLQQFQKEDA-
ALNIYEMLQNIFAIFRQDSSSTGWNETIVENLLANVYHQINHLKTVLEEKLEKEDFT-
RGKLMSSLHLKRYYGRILHYLKAKEYSHCAWTIVRVEILRNFYFINRLTGYLRN

REACTIVE SITES

.N.(13)		.D.(5)	.M.(3)	.Q.(11)	
4 YNL	86 ENL	34 KDR	36 RMN	10 LQR	51 FQK
14 SNF	90 ANV	39 FDI	62 EML	16 FQC	64 LQN
25 LNG	96 INH	54 EDA	117 LMS	18 CQK	72 RQD
37 MNF	153 RNF	73 QDS		23 WQL	94 HQI
58 LNI	158 INR	110 EDF		46 KQL	
65 QNI	166 RN			48 LQQ	
80 WNE				49 QQF	

HYDROFLEX PLOT



PREDICTED REACTIVITY AND DEGRADATION OF IFN-β

This protein has only a few sites predicted to be reactive, including an Asn-Gly (-LNG-) and three Mets. The Asn-Gly motif is located within a region predicted to be fairly hydrophilic but only moderately flexible. Stability studies of Ser-17-IFN-β at pH 5.5 with 0.1% SDS at 4°C for 1 year showed no detectable degradation of IFN-β by bioassay, SDS-PAGE, or RP-HPLC (Geigert *et al.*, 1988). Unfortunately no methods were used that were specific for detecting charged variants (such as an -LNG- to an -LDG- conversion), so it is uncertain whether the Asn-Gly residue reacts at 4°C. Oxidation of Met-62 was reported in liquid parenteral formulations under "conditions specific for Met oxidation" (Lin *et al.*, 1995). This protein was also found to be unstable in biological media (although the degradation pathway was not determined), indicative that the formulation and *ex vivo* stabilities may be different and depend on the nature of both the protein and the biological fluid (O'Kelley *et al.*, 1985).

Interferon-gamma (human) (γ-IFN) (144 residues)

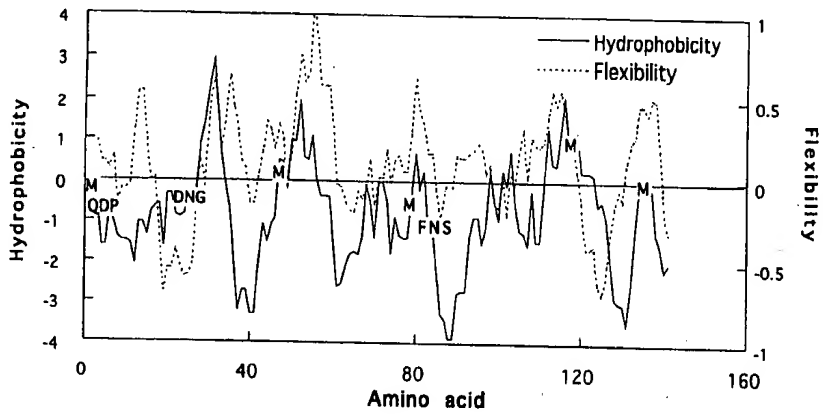
SEQUENCE

MQDPYVKEAENLKKYFNAGHSDVADNGTLFLGILKNWKEESDRKIMQSQIVSFYF-
KLFKNFKDDQSIQKSVETIKEDMNVKFFNSNKKKRDDFEKLTNYSVTDLNVQRKAI-
HELIQVMAELSPAAKTGKRKRSQMLFRGRRASQ

REACTIVE SITES

.N.(10)	.D.(10)	.M.(4)	.Q.(9)
11 ENL	3 QDP	46 IMQ	2 MQD
17 FNA	22 SDV	78 DMN	47 MQS
26 DNG	25 ADN	118 VMA	49 SQI
36 KNW	42 SDR	135 QML	65 DQS
60 KNF	63 KDD		68 IQK
79 MNV	64 DDQ		107 VQR
84 FNS	77 EDM		116 IQV
86 SNK	91 RDD		134 SQM
98. TNY	92 DDF		144 SQ
105 LNV	103 TDL		

HYDROFLEX PLOT



PREDICTED REACTIVITY AND DEGRADATION OF INTERFERON-GAMMA

γ-IFN has only a few sites predicted to undergo hydrolytic degradation (Asn-26, Asn-84). Asn-26 is predicted to be the faster of the two, based on primary amino acid sequence (Asn-26 is adjacent to Gly; Asn-84 is adjacent to Ser). Both Asn are found in regions of similar hydrophobicity, but Asn-26 is in a region of greater flexibility. It is not surprising that reaction at Asp was not observed, in that none of the Asp in γ-IFN are predicted hot spots based on primary sequence (no -XDG-). Similarly, there are nine Gln in γ-IFN, but again none are traditional hot spots (no -XQG-). The major degradation pathway of γ-IFN at neutral pH was found to be deamidation at Asn-26 and Asn-84 (Pearlman and Nguyen, 1992; Keck, 1995). A minor amount of Met oxidation was observed at Met-1 and Met-135. No evidence was found for cleavage at the -QDP- motif, nor reaction at Asp or Gln. At pH 5, the sum of these reaction rates did not compromise shelf life when the product is stored at 2–8°C. Interestingly, the covalent dimerization of γ-IFN has also been reported (Lauren *et al.*, 1993).

Interleukin-1 Receptor Antagonist (IL-1RA) (153 residues)

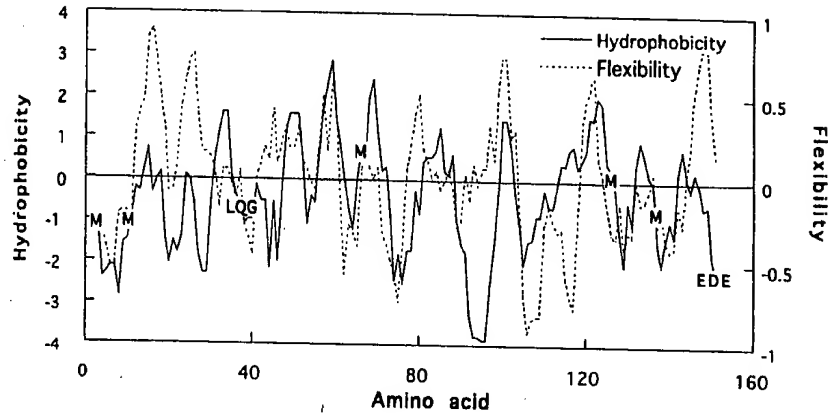
SEQUENCE

MRPSGRKSSKMQAFRIWDVNQKTFYLRNNQLVAGYLQGPNVNLEEKIDVVPIEPHA-
LFLGIHGGKMCLSCVKSGDETRLQLEAVNITDLSENRKQDKRFAFIRSDSGPTTSFES-
AACPGWFLCTAMEADQPVSLTNMPDEGVNVTKFYFQEDE

REACTIVE SITES

.N.(9)	.D.(9)	.M.(5)	.Q.(8)
20 VNQ	18 WDV	1 MR	12 MQA
28 RNN	48 IDV	11 KMQ	21 NQK
29 NNQ	75 GDE	66 KMC	30 NQL
40 PNV	88 TDL	126 AME	37 LQG
42 VNL	96 QDK	137 NMP	80 LQL
85 VNI	105 SDS		95 KQD
92 ENR	129 ADQ		130 DQP
136 TNM	139 PDE		150 FQE
143 VNV	152 EDE		

HYDROFLEX PLOT



PREDICTED REACTIVITY AND DEGRADATION OF IL-1RA

This protein has few reactive hot spots and is predicted to be fairly stable because the few reactive sites (Met and Gln-Gly) are often not shelf-life-limiting. The degradation of IL-1RA has been studied in some detail, and a number of unusual reaction sites have been observed (Maneri, 1994). IL-1RA formed a stable cyclic imide at Asp-152 (in the -EDE- motif) and underwent disulfide formation at Cys-68–Cys-71, and cyclization between the N- and C-termini of des-Glu-153. Minor degradation pathways were oxidation at the N-terminal Met (Met-1) and deamidation of Asn-136 (in the -TNM- motif). The pH of maximum stability was near pH 6, and aggregation was the primary route of degradation (and so not applicable to this analysis).

Interleukin-1α (IL-1α) (155 residues)

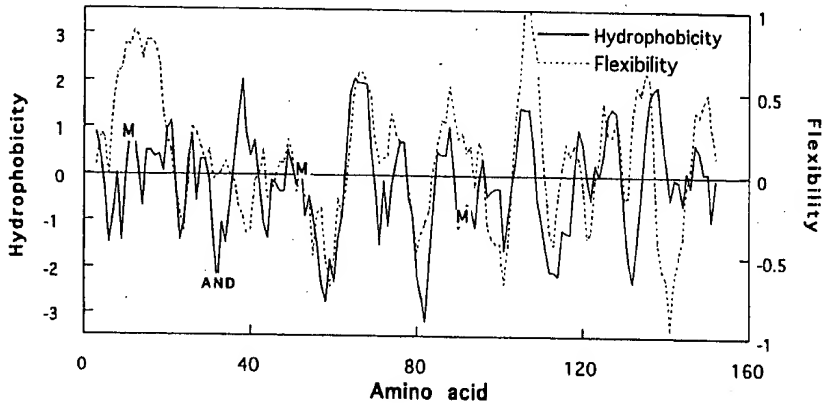
SEQUENCE

SFLSNVKYNFMRIIKYEFILNDALNQSIIRANDQYLTAAALHNLDEAVKFDMGAYKS-
SKDDAKITVILRISKTQLYVTAQDEDQPVLLKEMPEIPKTITGSETNLLFFWETHGTK-
NYFTSVAHPNLFIATKQDYWVCLAGGPPSITDFQILENQA

REACTIVE SITES

.N.(10)	.D.(10)	.M.(3)	.Q.(8)
5 SNV	22 DNA	11 FMR	26 NQS
9 YNF	33 NDQ	52 DMG	34 DQY
21 LND	45 LDE	91 EMP	74 TQL
25 LNQ	51 FDM		80 AQD
32 AND	60 KDD		84 DQP
43 HNL	61 DDA		132 KQD
104 TNL	81 QDE		149 FQI
116 KNY	83 EDQ		154 NQA
125 PNL	133 QDY		
153 ENQ	147 TDF		

HYDROFLEX PLOT



PREDICTED REACTIVITY AND DEGRADATION OF IL-1α

IL-1α lacks all of the traditional hydrolysis hot spots: Asn-Gly, Asn-Ser, Asp-Gly, and Gln-Gly. IL-1α contains three Met residues, and all of them are found in regions predicted to be fairly rigid, although it is not known if this should inhibit their oxidation. Based on the hydroflex plot, it is predicted that IL-1α should be fairly stable. The only residue found to degrade was Asn-32 (-AND-) giving a $t_{1/2}$ of approximately 25–30 hr at pH 7.5 and 42°C, considered accelerated reaction conditions. In retrospect, it is perhaps not surprising that reaction was observed at this site, in that it exists in a hydrophilic region of intermediate flexibility. Perhaps even more surprising is that this site is significantly more reactive than would be expected based on the reactivity found in small peptides (Robinson and Rudd, 1974). For example, the peptide GTND (where the TNX sequence is more reactive than the ANX motif) showed a half-life of 380 hr under the similar reaction conditions of pH 7.5 and 37°C, significantly longer than observed for the -AND- motif of IL-1α. The major degradation pathway of IL-1α at both pH 7.5 and pH 10.5 was found to be deamidation at Asn-32 (Wingfield *et al.*, 1987). This is in agreement with deamidation found in recombinant IL-1α purified from *E. coli*, as identified by the appearance of pI bands at 5.45 and 5.20, and by ^{1}H-NMR. Similar results were observed in another study describing the development of a deamidation-specific ELISA for detection of Asn-32-IL-1α and Asp-32-IL-1α (Sunahara *et al.*, 1989). No evidence was reported for Met oxidation. Unfortunately, the stability data presented in the literature were insufficient to determine whether or not this deamidation reaction proceeds sufficiently fast to compromise product shelf life at 2–8°C. IL-1α provides a clear cut example showing that the prediction of protein reactivity is not straightforward, when basing the prediction on a primary sequence analysis and reactivity in small peptides of similar motif. It is satisfying to find that Asn-32 is harbored in a hydrophilic region of intermediate flexibility. IL-1α also shows that the -XND- motif can be sufficiently reactive in aqueous solution that it may become the dominant site of degradation if other, more reactive sites (such as Asn-Gly, Asn-Ser, Asp-Gly, and Gln-Gly) are not available.

Interleukin-1β (human) (IL-1β) (153 residues)

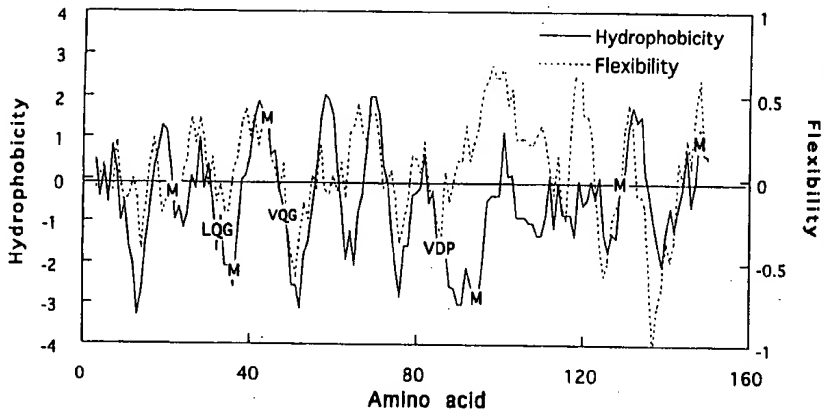
SEQUENCE

APVRSLNCTLRDSQQKSLVMSGPYELKALHLQGQDMEQQVVFSMSFVQGEESNDK-
IPVALGLKEKNLYLSCVLKDDKPTLQLESVDPKNYPKKKMEKRFVFNKIEINNKLEF-
ESAQFPNWYISTSQAENMPVFLGGTKGGQDITDFTMQFVSS

REACTIVE SITES

.N.(9)	.D.(8)	.M.(6)	.Q.(12)	
7 LNC	12 RDS	20 VMS	14 SQQ	116 AQF
53 SND	35 QDM	36 DME	15 QQK	126 SQA
66 KNL	54 NDK	44 SMS	32 LQG	141 GQD
89 KNY	75 KDD	95 KME	34 GQD	149 MQF
102 FNK	76 DDK	130 NMP	38 EQQ	
107 INN	86 VDP	148 TMQ	39 QQV	
108 NNK	142 QDI		48 VQG	
119 PNW	145 TDF		81 LQL	
129 ENM				

HYDROFLEX PLOT



PREDICTED REACTIVITY AND DEGRADATION OF IL-1β

Inspection of the human sequence suggests that the most likely site for deamidation is either Gln-32 or Gln-48. Although these Gln are preceded by bulky hydrophobic residues (-LQG- and -VQG) that are deactivating, both of these Gln are in regions of moderate hydrophobicity and flexibility which may allow their reaction. An alternative, but less likely, reaction site is Asn-53 (-SND-), in that it is activated by the preceding Ser and exists in a region predicted to be hydrophilic and flexible. The major degradation pathway of IL-1β at neutral pH and temperatures less than 30°C was reported to be deamidation, although the site of deamidation was not determined (Gu *et al.*, 1991). It was reported that murine recombinant IL-1β selectively deamidated at Asn-32 (-LNG-), but this sequence is not found in human IL-1β, as it is modified to contain Gln (-LQG-) of lower chemical reactivity (Daumy *et al.*, 1991). Modification at this site did not lead to complete inactivity, wherein the deamidated product had ~50% of the original activity. The reactivity of IL-1β was sufficiently slow at temperatures less than 5°C that it was predicted that this reaction would not compromise the formulation shelf life. H_2O_2-catalyzed oxidation has been observed at Met-20, Met-36 or -44, Met-130, and Met-148 (Foster, 1996).

- **Interleukin-1β (murine) (152 residues)**

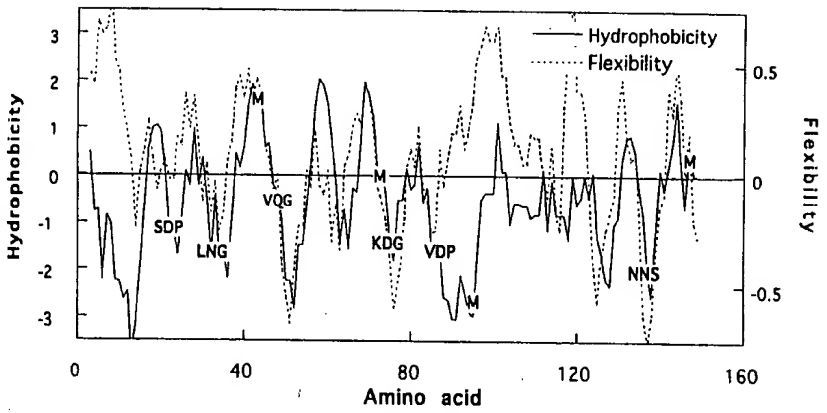
SEQUENCE

VPIRQLHYRLRDEQQKSLVLSDPYELKALHLNGQNINQQVIFSMSFVQGEPSNDKIP-
VALGLKGKNLYLSCVMKDGTPTLQLESVDPKQYPKKKMEKRFVFNKIEVKSKVEF-
ESAEFPNWYISTSQAEHKPVFLGNNSGQDIIDFTMESVSS

REACTIVE SITES

.N.(9)	.D.(7)	.M.(4)	.Q.(11)
32 LNG	12 RDE	44 SMS	5 RQL
35 QNI	22 SDP	73 VMK	14 EQQ
37 INQ	54 NDK	95 KME	15 QQK
53 SND	75 KDG	147 TME	34 GQN
66 KNL	86 VDP		38 NQQ
102 FNK	141 QDI		39 QQV
119 PNW	144 IDF		48 VQG
136 GNN			81 LQL
137 NNS			89 KQY
			126 SQA
			140 GQD

HYDROFLEX PLOT



PREDICTED REACTIVITY AND DEGRADATION OF IL-1β (MURINE)

Inspection of the hydroflex plot shows that the -LNG- motif (Asn-32) lies in a hydrophilic region of moderate flexibility and is the most likely reaction site. Interestingly, Asn-137 (-NNS-) also resides in a hydrophilic flexible region and may also react slightly. Comparison of the amino acid sequences for human and murine IL-1β show that neither of these predicted hot spots is available for reaction on human IL-1β, so it is expected that murine and human IL-1β should have different degradation pathways (see comparison of sequences in Scheme 11). Incubation of murine IL-1β in pH 8.5 aqueous solution at 37°C for 35 hr afforded deamidated IL-1β, where deamidation occurred primarily at Asn-32 (original numbering, Asn-149). Although the tryptic maps on IL-1β (murine) were inconclusive for deamidation in several regions of the molecule, they did show that the C-terminal end (containing the -NNS- motif)

Interleukin-11 (human) (178 residues)

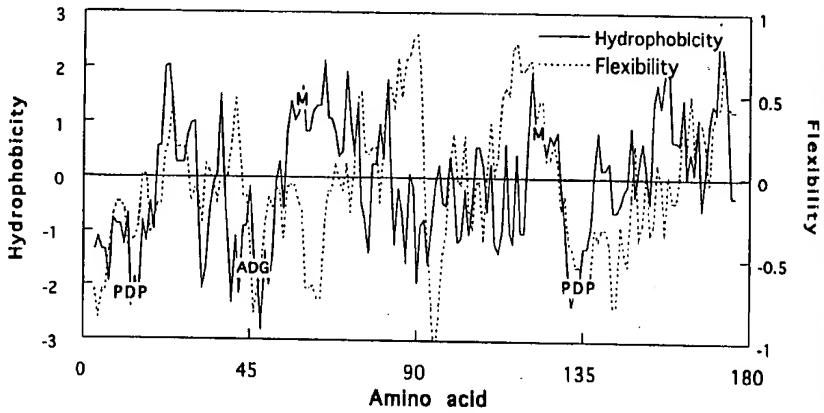
SEQUENCE

PGPPPGPPRVSPDPRAELDSTVLLTRSLLADTRQLAAQLRDKFPADGDHNLDSLPTL-
AMSAGALGALQLPGVLTRLRADLLSYLRHVQWLRRAGGSSLKTLEPELGTLQARL-
DRLLRRLQLLMSRLALPQPPPDPPAPPLAPPSSAWGGIRAAHAILGGLHLTLDWAVR-
GLLLLKTRL

REACTIVE SITES

.N.(1)	.D.(11)	.M.(2)	.Q.(7)
50 HNL	13 PDP	59 AMS	34 RQL
	19 LDS	123 LMS	38 AQL
	31 ADT		68 LQL
	41 RDK		88 VQW
	46 ADG		109 LQA
	48 GDH		120 LQL
	52 LDS		130 PQP
	79 ADL		
	113 LDR		
	134 PDP		
	165 LDW		

HYDROFLEX PLOT



PREDICTED REACTIVITY AND DEGRADATION OF INTERLEUKIN-11

The hydroflex plot for IL-11 shows that there are only a few predicted reactive sites for degradation. These include two Asp-Pro linkages that may be susceptible to acid-catalyzed cleavage: Asp-Gly, which may form iso-Asp-Gly, and Met oxidation. An excellent study on IL-11 degradation by Ingram and Warne (1994) included the effect of pH on the different pathways and degradation due to dimerization and aggregation. Briefly, IL-11 showed cleavage between Asp-13–Pro-14 and Asp-134–Pro-135 under acid conditions, but only minor amounts of cleavage at pH 7.2 after 146 days. IL-11 also showed some deamidation at higher pHs at Asn-50 (6.2%/wk at pH 9.6 and 30°C), but only minor amounts at lower pHs (0.19%/wk at pH 5.5 at 30°C). Based on these rates, it is roughly predicted that deamidation at these sites would not compromise the shelf life at 2–8°C. Some oxidation of Met-59 was also observed, especially at lower pH, but was minor at neutral pH for most of the buffers studied. No degradation of the Asp-Gly site was reported (-ADG-), although it is unknown whether the analytical methods used would have detected this.

Lung Surfactant SP-C (human) (34 residues)

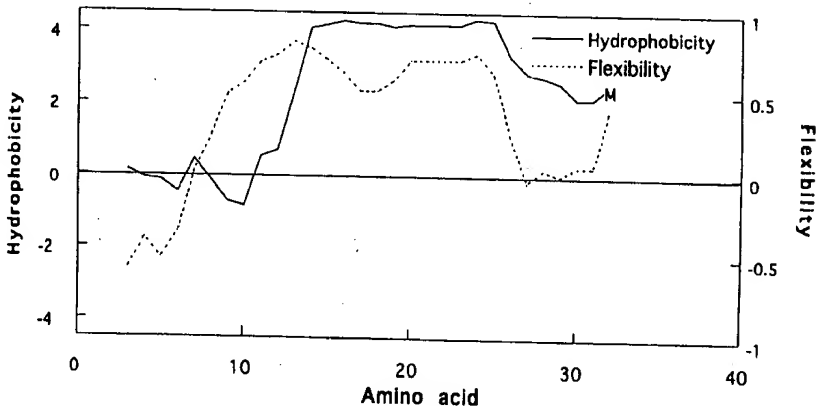
SEQUENCE

GIPSSPVHLKRLLIVVVVVVLIVVVIVGALLMGL

REACTIVE SITES

.N.(0)	.D.(0)	.M.(1)	.Q.(0)
		32 LMG	

HYDROFLEX PLOT



PREDICTED REACTIVITY AND DEGRADATION OF LUNG SURFACTANT

Recombinant human lung surfactant is predicted to be extremely resistant to hydrolytic degradation, notably because of a complete absence of reactive sites. Lung surfactant does have a single Met near its C-termini, although this is in a hydrophobic region that is predicted to be fairly inflexible. Lung surfactant did not show any hydrolysis after reconstitution and storage at 2–8°C after 1 month. This polypeptide was susceptible to oxidation at Met-32, the only site of predominant chemical reactivity.

Lysozyme (hen egg white) (129 residues)

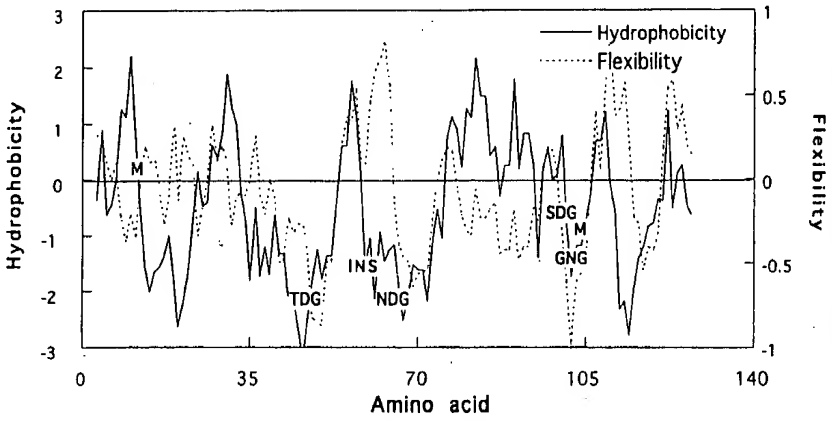
SEQUENCE

KVFGRCELAAAMKRHGLDNYRGYSLGNWVCAAKFESNFNTQATNRNTDGSTDYG-
ILQINSRWWCNDGRTPGSRNLCNIPCSALLSSDITASVNCAKKIVSDGNGMNAWVA-
WRNRCKGTDVQAWIRGCRL

REACTIVE SITES

.N.(14)	.D.(7)	.M.(2)	.Q.(3)
19 DNY	65 CND	18 LDN	12 AMK	41 TQA
27 GNW	74 RNL	48 TDG	105 GMN	57 LQI
37 SNF	77 CNI	52 TDY		121 VQA
39 FNT	93 VNC	66 NDG		
44 TNR	103 GNG	87 SDI		
46 RNT	106 MNA	101 SDG		
59 INS	113 RNR	119 TDV		

HYDROFLEX PLOT



PREDICTED REACTIVITY AND DEGRADATION OF LYSOZYME

Inspection of the primary amino acid sequence for lysozyme shows that there are several reactive residues, including Asn-103 within the -GNG- motif, and several Asp-Gly residues. Based on this, it is anticipated that lysozyme should be fairly unstable in aqueous solution; indeed, anecdotal observations of lysozyme instability likely prompted the seminal peptide model studies of Robinson and colleagues (Robinson and Tedro, 1973b). Unfortunately, the degradation pathways of lysozyme itself were not studied in depth nor monitored using chromatographic techniques where subtle changes such as Asp conversion to iso-Asp would be detected. It was noted that the primary amino acid sequence consisted of Gly-Asp-Gly instead of Gly-Asn-Gly at positions 102–104 (Canfield, 1963), indicating a high propensity for deamidation at this reactive site. There is also another report showing cyclic imide formation of Asp-101 in the -SDG- motif when incubated at 40°C and pH 4 (Tomizawa *et al.*, 1994). This motif is known to be located within a solvent-accessible and flexible region. Interestingly, the -TDG- motif is also located in a region of predicted hydrophilicity and flexibility, and yet the authors did not report cyclic imide formation or iso-Asp formation at this site. No account of lysozyme oxidation was reported in aqueous formulation.

Myelin Basic Protein (MBP) (169 residues)

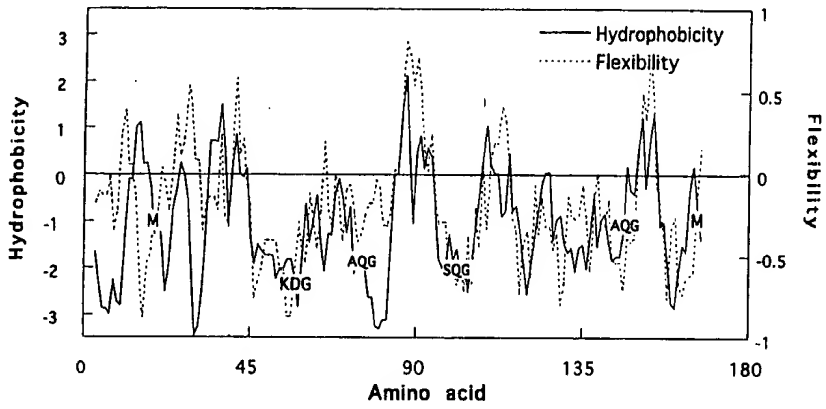
SEQUENCE

AAQKRPSQRSKYLASASTMDHARHGFLPRHRDTGILDSLGRFFGSDRGAPKRGSGK-
DGHHAARTTHYGSLPQKAQGHRPQDENPVVHFFKNIVTPRTPPPSQGKGRGRSLSR-
FSWGAEGQKPGFGYGGRASDYKSAHKGLKGHDAQGTLSKIFKLGGRDSRSGSPM-
ARR

REACTIVE SITES

.N.(2)	.D.(9)	.M.(2)	.Q.(8)
83 ENP	20 MDH	19 TMD	3 AQK
91 KNI	32 RDT	166 PMA	8 SQR
	37 LDS		72 PQK
	46 SDR		75 AQG
	57 KDG		80 PQD
	81 QDE		102 SQG
	132 SDY		120 GQK
	144 HDA		146 AQG
	159 RDS		

HYDROFLEX PLOT



PREDICTED REACTIVITY AND DEGRADATION OF MYELIN BASIC PROTEIN

MBP has a single Asp-Gly (-KDG-) that may undergo cyclization and iso-Asp formation, as well as some Gln-Gly residues or predicted lesser reactivity. Both of these Gln are located in hydrophilic regions of modest flexibility. Isolation of bovine MBP resulted in partial deamidation of the Gln residues at positions 102 and 146 (corrected numbering) (Chou *et al.*, 1976). Unfortunately, it was not possible to distinguish whether the microheterogeneity was present in the original protein or was due to the work-up carried out at pH 10.4. There also exists a reactive Asp-Gly linkage, but it is unlikely that the paper chromatography methods used in this paper would isolate the iso-Asp product.

Neocarzinostatin (109 residues)

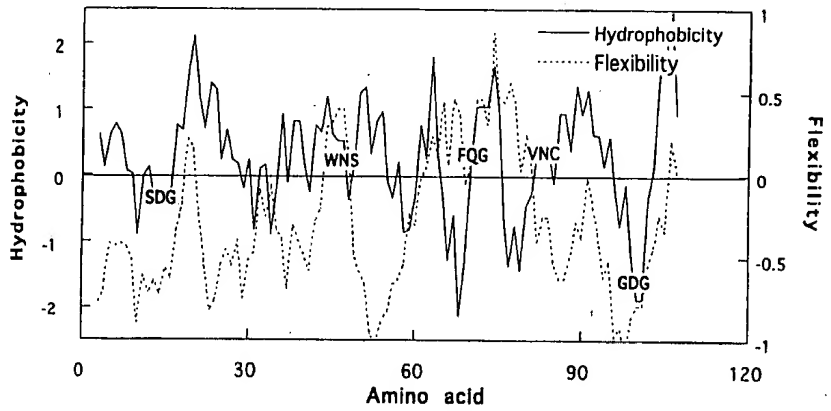
SEQUENCE

AAPTATVTPSSGLSDGTVVKVAGAGLQAGTAYDVGQCASVNTGVLWNSVTAAGSA-CDPANFSLTVRRSFQGFLFDFTRWGTVNCTTAACQVGLSDAAGDGQPGVAISFN

REACTIVE SITES

.N.(5)		.D.(6)		.M.(0)	.Q.(5)	
41 VNT	83 VNC	15 SDG	75 FDF		27 LQA	90 CQV
47 WNS	109 FN	33 YDV	95 SDA		36 GQC	101 GQP
60 ANF		57 CDP	99 GDG		70 FQG	

HYDROFLEX PLOT



PREDICTED REACTIVITY AND DEGRADATION OF NEOCARZINOSTATIN

Inspection of the primary amino acid sequence for neocarzinostatin shows that there are several reactive hydrolysis sites, including Asn-47 (-WNS-), Asp-15 (-SDG-), and Asp-99 (-GDG-). The Asn-47 site may be only mildly reactive, in that it is in a region of moderate hydrophobicity. On the other hand, Asp-99 is in a flexible, hydrophilic region and is expected to be reactive. Under weakly acidic conditions at 4°C the major degradation pathway of neocarzinostatin was conversion of Asn-83 to Asp-83 (Maeda and Kuromizu, 1977). No other degradation products were observed during the several-day course of the reaction. Because these experiments were carried out at pH 3.2 (somewhat lower than would likely be used in a protein parenteral liquid formation), the rate data obtained in this paper are of limited utility in determining the preferred pathway at intermediate pH or for estimating whether or not neocarzinostatin would exhibit a shelf life of 2 years between pH 5 and pH 7. This protein shows that reaction in aqueous solution may occur at sites other than the traditional hot spots (or that pH is crucial in making predictions and that data at pH 3 should not be used to predict the major degradation pathways at pH 5–7). From the methods used, it is unlikely that iso-Asp formation would have been detected at Asp-15 or Asp-99, so it is unknown whether or not reaction at these hot spots actually occurred.

Nerve Growth Factor (human) (NGF) (120 residues)

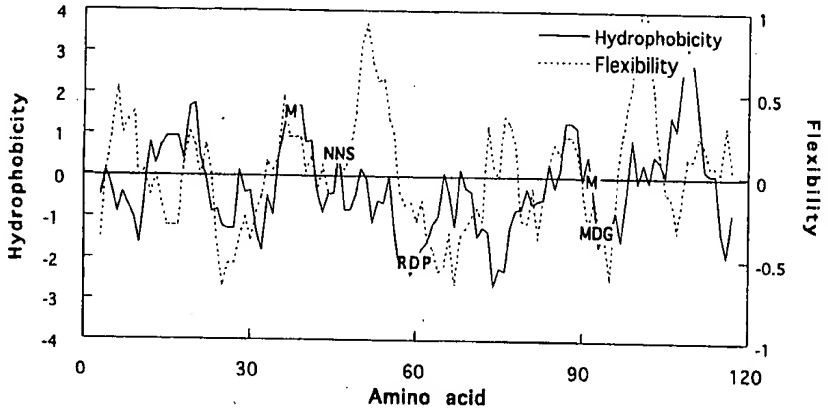
SEQUENCE

SSSHPIFHRGEFSVCDSVSVWVGDKTTATDIKGKEVMVLGEVNINNSVFKQYFFETK-CRDPNPVDSGCRGIDSKHWNSYCTTTHTFVKALTMDGKQAAWRFIRIDTACVCVLS-RKAVRRA

REACTIVE SITES

.N.(5)	.D.(8)	.M.(2)	.Q.(2)
43 VNI	16 CDS	37 VMV	51 KQY
45 INN	24 GDK	92 TMD	96 KQA
46 NNS	30 TDI		
62 PNP	60 RDP		
77 WNS	65 VDS		
	72 IDS		
	93 MDG		
	105 IDT		

HYDROFLEX PLOT



PREDICTED REACTIVITY AND DEGRADATION OF NGF

The primary sequence of NGF indicated that there are three hot spots for hydrolytic degradation: Asn-46 (-NNS-), Asp-93 (-MDG-), and Asp-60 (-RDP-). This last motif is sensitive to acid-catalyzed cleavage only, and is generally stable above pH 5 at 2–8°C. Of the other two motifs, Asp-93 resides in a hydrophilic region that is calculated to be fairly flexible, and so this site might be expected to be reactive. Asn-46 is adjacent to the Ser and is expected to be only moderately activated (as compared to Gly). It resides in a region of intermediate hydrophobicity and flexibility, and so may be expected to be only moderately reactive. Met-37 resides in a hydrophobic, inflexible region. It was shown that the primary degradation site in NGF at pH 5.5 was iso-Asp formation at Asp-93 (-MDG-). Only minor amounts of deamidation were observed at Asn-45 (-INN-), a site not predicted to be normally reactive, and it was believed that this may have occurred in the processing steps at higher pH. Also minor amounts of Met oxidation were found, both at Met-37 and Met-92. The sum of all these hydrolytic and oxidative degradation reactions did not compromise the shelf life of liquid parenteral formulations of NGF at 2–8°C.

Parathyroid Hormone (84 residues)

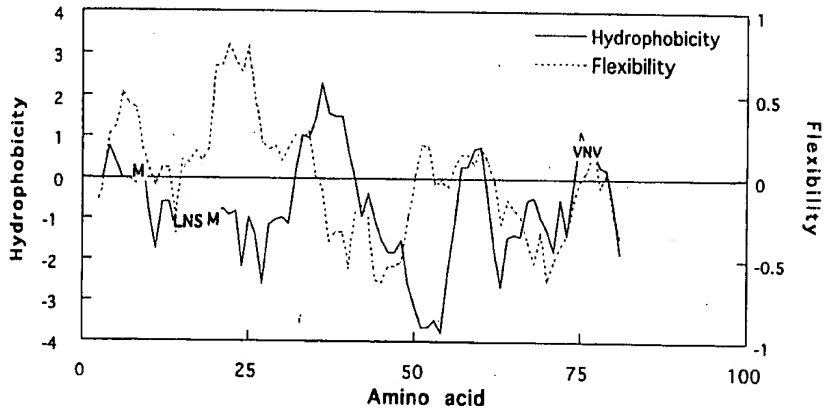
SEQUENCE

SVSEIQLMHNLGKHLNSMERVEWLRKKLQDVHNFVALGAPLAPRDAGSQRPRKKE-DNVLVESHEKSLGEADKADVNVLTKAKSQ

REACTIVE SITES

.N.(5)	.D.(5)	.M.(2)	.Q.(4)
10 HNL	30 QDV	8 LMH	6 IQL
16 LNS	45 RDA	18 SME	29 LQD
33 HNF	56 EDN		49 SQR
57 DNV	71 ADK		84 SQ
76 VNV	74 ADV		

HYDROFLEX PLOT



PREDICTED REACTIVITY AND DEGRADATION OF PARATHYROID HORMONE

This protein has only a few hot spots, including Asn-16 in the -LNS- motif and the Met sites. For years it was believed that parathyroid hormone contained Asp at position 76, in that all reports of extracted and purified human, bovine, or porcine parathyroid hormone contained Asp-76 (Keutmann *et al.*, 1978). Later, however, nucleotide sequencing of cloned cDNAs encoding human parathyroid hormone messenger RNA showed that the correct residue was Asn-76 (Hendy *et al.*, 1981). This is another dramatic example of *in vivo* deamidation and so complete that no microheterogeneity was observed at position 76. This reaction occurred at an unlikely site, that is, within the -VNV- motif. Further, this reactive site is in a region that is not particularly hydrophilic or flexible. No control experiments were carried out to show that the

same deamidation reaction occurs in pH 4.5–7.5 formulation buffer. The oxidation of parathyroid hormone has also been studied using H_2O_2 as catalyst. After incubation at pH 10 with 1 mM H_2O_2, both Met-8 and Met-18 were oxidized, giving the two monooxidized products, as well as the dioxidized product. Biological activity was reduced moreso after oxidation at Met-8 than Met-18 (Nabuchi *et al.*, 1995).

Relaxin

SEQUENCE (A CHAIN) (24 residues)

QLYSALANKCCHVGCTKRSLARFC

SEQUENCE (B CHAIN) (29 residues)

DSWMEEVIKLCGRELVRAQIAICGMSTWS

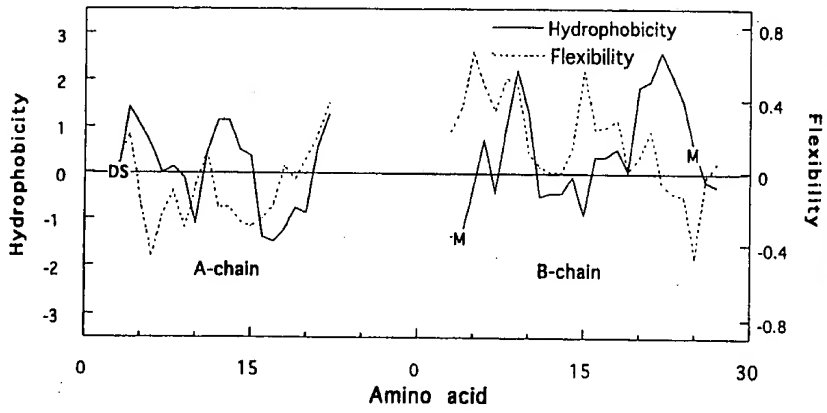
REACTIVE SITES (A CHAIN)

.N.(1)	.D.(0)	.M.(0)	.Q.(0)
8 ANK			

REACTIVE SITES (B CHAIN)

.N.(0)	.D.(1)	.M.(2)	.Q.(1)
	1 DS	4 WME	19 AQI
		25 GMS	

HYDROFLEX PLOT



PREDICTED REACTIVITY AND DEGRADATION OF RELAXIN

This small protein has few predicted reactive sites, perhaps the likeliest being oxidation of the Met residues on the relaxin B chain. It was shown that the predominant cleavage pathway for relaxin at pH 3–5 was cleavage of the N-terminal Asp on the B chain (Nguyen *et al.*, 1993a). At pH 5–7, the major degradation pathways were again cleavage of this Asp, and oxidation of Met-4 and Met-25 on the B chain (Cipolla and Shire, 1991), in agreement with hydrogen peroxide-catalyzed oxidation studies (Nguyen *et al.*, 1993a). Disulfide scrambling occurred at higher pHs (Canova-Davis *et al.*, 1990, 1991).

Ribonuclease A (RNase A) (124 residues)

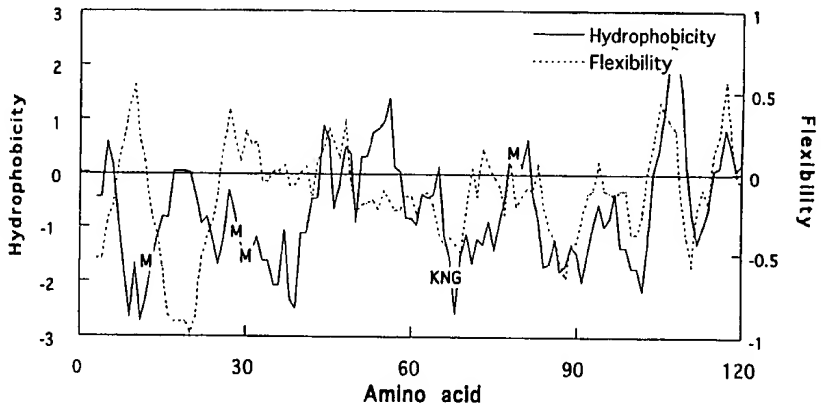
SEQUENCE

KETAAAKFERQHMDSSTSAASSSNYCNQMMKSRNLTKDRCKPVNTFVHESLADV-
QAVCSQKNVACKNGQTNCYQSYSTMSITDCRETGSSKYPNCAYKTTQANKHIIVAC-
EGNPYVPVHFDASV

REACTIVE SITES

.N.(10)		.D.(5)	.M.(4)	.Q.(7)	
24 SNY	67 KNG	14 MDS	13 HMD	11 RQH	74 YQS
27 CNQ	71 TNC	38 KDR	29 QMM	28 NQM	101 TQA
34 RNL	94 PNC	53 ADV	30 MMK	55 VQA	
44 VNT	103 ANK	83 TDC	79 TMS	60 SQK	
62 KNV	113 GNP	121 FDA		69 GQT	

HYDROFLEX PLOT



PREDICTED REACTIVITY AND DEGRADATION OF RNase A

Inspection of the primary amino acid sequence, and the hydropathy and flexibility plots for RNase A shows that there is a single site for facile hydrolytic degradation at Asn-67 in the -KNG- motif. Based on this, Asn-67 is the most likely site of degradation. There are also several Met residues found in fairly hydrophilic regions of varying flexibility. This protein degraded primarily at its predicted hot spot (Asn-67) at both low and high pH. Reaction under strong acid conditions at 30°C showed reaction at Asn-67 (Venkatesh and Vithayathil, 1984) as it did at pH 8 and above (Bornstein and Balian, 1970; Wearne and Creighton, 1989). The conformation of this protein played a major role in its rate of deamidation, as shown by deamidation studies of ribonuclease (Bornstein and Balian, 1970; Wearne and Creighton, 1989) where an Asn that ordinarily does not deamidate in the native structure deamidates in the denatured protein. The oxidative degradation pathways of RNase A have not been reported.

Ribonuclease U2 (RNase U2) (*Ustilago sphaerogena*) (114 residues)

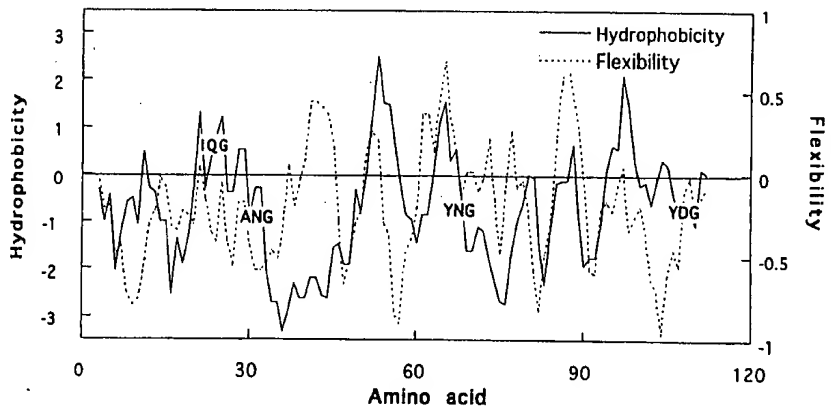
SEQUENCE

CDIPQSTNCGGNVYSNDDINTAIQGALDDVANGDRPDNYPHQYYDEASDQITLCC-GSGPWSEFPLVYNGPYYSSRDNYVSPGPDRVIYQTNTGEFCATVTHTGAASYDGFT-QCS

REACTIVE SITES

.N.(9)	.D.(12)	.M.(0)	.Q.(6)
8 TNC	2 CDI		5 PQS
12 GNV	17 NDD		24 IQG
16 SND	18 DDI		42 HQY
20 INT	28 LDD		50 DQI
32 ANG	29 DDV		89 YQT
38 DNY	34 GDR		112 TQC
68 YNG	37 PDN		
77 DNY	45 YDE		
91 TNT	49 SDQ		
	76 RDN		
	84 PDR		
	108 YDG		

HYDROFLEX PLOT



PREDICTED REACTIVITY AND DEGRADATION OF RNase U2

Inspection of the primary amino acid sequence and the hydropathy and flexibility plots for ribonuclease U2 shows that there is a site for facile hydrolytic degradation at Asn-32 in the -ANG- motif and another Asn-Gly in the -YNG- motif. There also exists an Asp-Gly near the C-termini that may be susceptible to iso-Asp formation. Isolation of this protein results in two isoforms, RNase U2-A and RNase U2-B, of which the major difference in these is a change in the protein pI (Kanaya and Uchida, 1986). Degradation of this protein at one of its predicted hot spots (Asn-32) is the cause of the RNase U2-B isoform, where an iso-Asp-Gly linkage was found. The catalyst for this deamidation reaction was not determined but was likely due to simple pH catalysis during work-up (although enzymatic catalysis during fermentation cannot be ruled out based on the conditions used). Of interest, there is also another Asn-Gly motif in RNase U2 (-YNG-) that is surprisingly resistant to hydrolytic modification. Asn-68 is in a region of similar predicted hydrophobicity, and only slightly less flexible than Asn-32, and remained stable as the Asn form.

Secretin (27 residues)

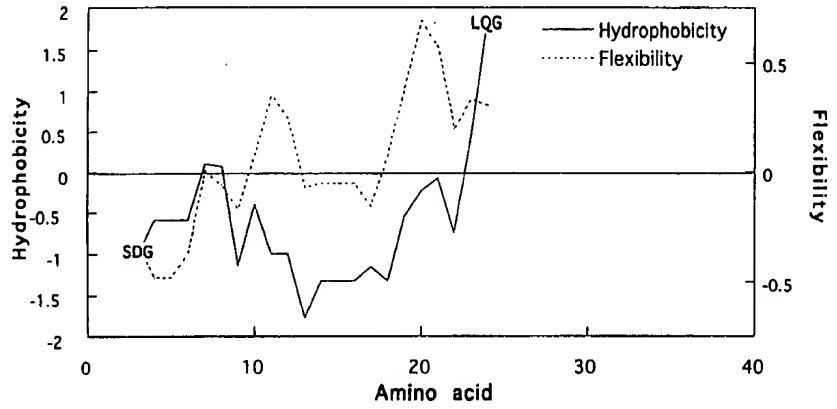
SEQUENCE

HSDGTFTSELSRLRDSARLQRLLQGLV

REACTIVE SITES

.N.(0)	.D.(2)	.M.(0)	.Q.(2)
	3 SDG		20 LQR

HYDROFLEX PLOT



PREDICTED REACTIVITY AND DEGRADATION OF SECRETIN

The amino acid sequence for secretin is short, with a concomitant few number of reactive sites. Nevertheless, the predicted site of reactivity is reaction at Asp-3 (-SDG-), and this should predominate easily over reaction at Gln-24 in the -LQG- motif. The -SDG- motif is also found in a fairly hydrophilic and flexible environment, suggesting that reaction may be possible (this flexibility calculation may be of little value in a peptide of this size which is likely to be highly flexible through its entire length). The major degradation pathway of secretin at neutral pH was reaction at Asp-3 to give the iso-Asp-3 product (Tsuda *et al.*, 1990). In this study, the degradation of secretin was carried out at 60°C, much higher than the 2–30°C likely for storage of an aqueous secretin formulation. This example is still included, however, in that secretin is a small peptide rather than a protein, and data obtained at higher temperatures under accelerated stability conditions are likely to mimic the reaction pathway observed at lower temperatures. Unfortunately, these authors did not carry out their stability kinetics at different temperatures, so it is not possible to estimate the shelf life for secretin at 2–8°C.

Serine Hydroxymethyltransferase (SHMT) (rabbit) (483 residues)

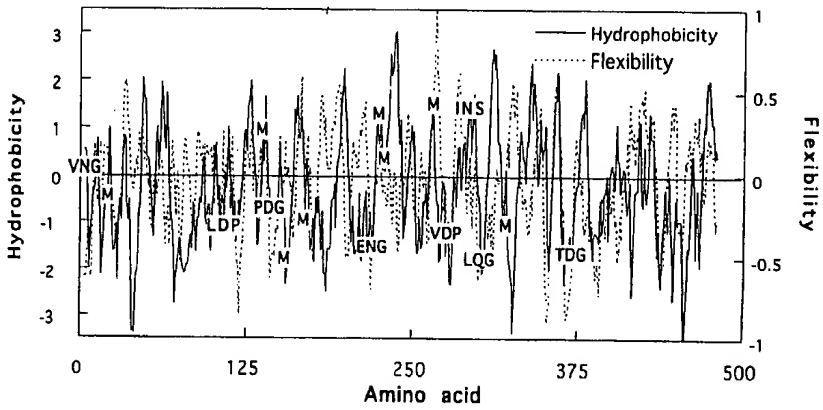
SEQUENCE

ATAVNGAPRDAALWSSHEQMLAQPLKDSDAEVYDIIKKESNRQRVGLELIASENFAS-
RAVLEALGSCLNNKYSEGYPGQRYYGGTEHIDELETLCQKRALQAYGLDPQCWGV-
NVQPYSGSPANFAVYTALVEPHGRIMGLDLPDGGHLTHGFMTDKKKISATSIFFESM-
AYKVNPDTGYIDYDRLEENARLFHPKLIIAGTSCYSRNLDYGRLRKIADENGAYLM-
ADMAHISGLVVAGVVPSPFEHCHVVTTTTHKTLRGCRAGMIFYRRGVRSVDPKTGK-
EILYNLESLINSAVFPGLQGGPHNHAIAGVAVALKQAMTPEFKEYQRQVVANCRAL-
SAALVELGYKIVTGGSDNHLILVDLRSKGTDGGRAEKVLEACSIACNKNTCPGDKS-

REACTIVE SITES

.N.(18)	.D.(24)	.M.(8)	.Q.(17)
5 VNG	10 RDA	20 QML	19 EQM
41 SNR	27 KDS	138 IMG	23 AQP
54 ENF	29 SDA	153 FMT	43 RQR
69 LNN	34 YDI	169 SMA	79 GQR
70 NNK	89 IDE	225 LMA	96 CQK
113 VNV	106 LDP	228 DMA	101 LQA
123 ANF	141 LDL	265 GMI	108 PQC
174 VNP	144 PDG	319 AMT	115 VQP
188 ENA	155 TDK		300 LQG
207 RNL	176 PDT		317 KQA
220 ENG	181 IDY		327 YQR
286 YNL	183 YDR		329 RQV
292 INS	209 LDY		418 FQK
305 HNH	218 ADE		433 VQI
333 ANC	227 ADM		435 IQD
355 DNH	276 VDP		458 HQR
384 CNK	354 SDN		466 RQE
386 KNT	361 VDL		
	368 TDG		
	391 GDK		
	416 KDF		
	436 QDD		
	437 DDT		
	454 GDE		

HYDROFLEX PLOT



PREDICTED REACTIVITY AND DEGRADATION OF SHMT

expected to show deamidation. The second, Asn-220, is found in a motif predicted to be fairly reactive (-ENG-) and is found in a hydrophilic flexible region. Other reactive hot spots include Asp-144 (-PDG-), Asp-368 (-TDG-), and Gln-300 (-LQG-), all of which are found in fairly hydrophilic, flexible regions of the protein. Artigues *et al.* (1990) demonstrated that SHMT deamidated *in vivo* at Asn-5 to give iso-Asp-5. In addition, they carried out a short set of control experiments and showed that this deamidation reaction was not a consequence of the purification work-up, and that deamidation occurred at pH 7.3 and 37°C. They found that the Asn-5 moiety in SHMT disappeared with a half-life of 450 hr, significantly slower than model peptides Ac-VNGA ($t_{1/2}$ = 80 hr) and Ac-ATAVNGAPRDAALW ($t_{1/2}$ = 70 hr) of the identical N-terminal sequence. The work-up procedure (chymotryptic cleavage of the N-terminal 15-mer followed by HPLC analysis) precluded determination of deamidation at other sites in the protein. No analysis was made to determine if Met oxidation occurred upon storage in aqueous solution. The degradation rate constants were determined at only 37°C, so no extrapolation can be made as to the stability at 2–8°C.

Tissue Factor-243 (243 residues)

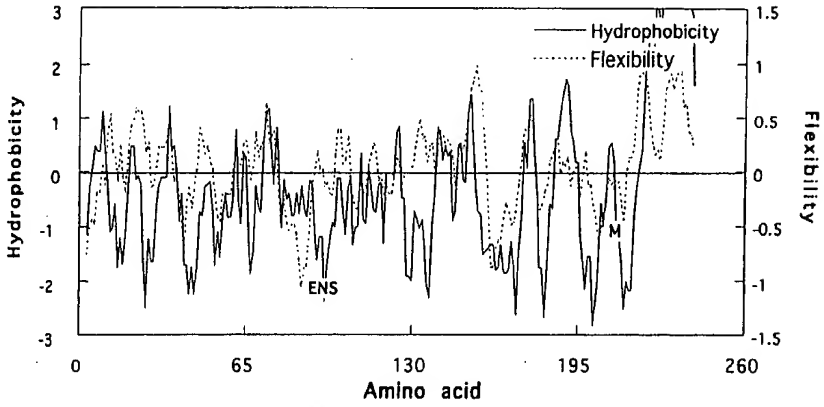
SEQUENCE

SGTTNTVAAYNLTWKSTNFKTILEWEPKPVNQVYTVQISTKSGDWKSKCFYTTDTE-
CDLTDEIVKDVKQTYLARVFSYPAGNVESTGSAGEPLYENSPEFTPYLETNLGQPTIQ-
SFEQVGTKVNVTVEDERTLVRRNNTFLSLRDVFGKDLIYTLYYWKSSSSGKKTAKT-
NTNEFLIDVDKGENYCFSVQAVIPSRTVNRKSTDSPVECMGQEKGEFREIFYIIGAVV-
FVVIILVIILAISLH

REACTIVE SITES

.N.(14)	.D.(11)	.M.(1)	.Q.(8)
5 TNT	44 GDW	210 CMG	32 NQV
11 YNL	54 TDT		37 VQI
18 TNF	58 CDL		69 KQT
31 VNQ	61 TDE		110 GQP
82 GNV	66 KDV		114 IQS
96 ENS	129 EDE		118 EQV
107 TNL	145 RDV		190 VQA
124 VNV	150 KDL		212 GQE
137 RNN	178 IDV		
138 NNT	180 VDK		
171 TNT	204 TDS		
173 TNE			
184 ENY			
199 VNR			

HYDROFLEX PLOT



PREDICTED REACTIVITY AND DEGRADATION OF TISSUE FACTOR-243

Tissue factor is a blood coagulation protein cofactor which exists as a glycosylated integral membrane protein. A truncated form of tissue factor that includes the transmembrane domain (amino acids 1–243) has been developed, and some stability data exist for this truncated form. Inspection of the amino acid sequence of TF-243 shows that, even though there are a large number of Asn, Asp, and Gln, only one residue (Asn-96) in the -ENS- motif is a predicted hot spot. This motif containing Ser rather than Gly adjacent to Asn is predicted to be only moderately reactive. Note, however, that Asn-96 resides in a hydrophilic region, but of only intermediate flexibility. Formulation of the truncated form of tissue factor at 0.1 mg/ml and pH 7.3 in 10 mM isotonic pH 7.3 sodium phosphate and 0.8% octylglucoside showed no signs of degradation by several different methods when stored for 0.5 year at 2–8°C (Shire, 1995). No alterations in tissue factor were detected by SDS PAGE, size-exclusion chromatography, ELISA, chromogenic and clotting activity assays after 54 weeks at 2–8°C and at 25°C, when compared to a sample stored at −70°C. In the starting material stored at −70°C there were two bands at approximately pI 5.3, and after 54 weeks at 2–8°C another band was formed at ~pI of 5.2. At 25°C, the band at pI 5.2 was more intense than the doublet of bands at pI 5.3, and there was also an additional band at pI 5.0, whereas the original doublet at pI 5.3 was barely visible. The generation of acidic components suggests that deamidation occurred during storage, but was not conclusively proven. The activity of tissue factor as determined by a chromogenic assay remained unaltered after 54 weeks at 2–8°C but decreased by 33% during storage at 25°C. Extrapolation of these data suggest that TF-243 is sufficiently stable to permit storage at 2–8°C for at least 18 months.

TGF-Beta (112 residues)

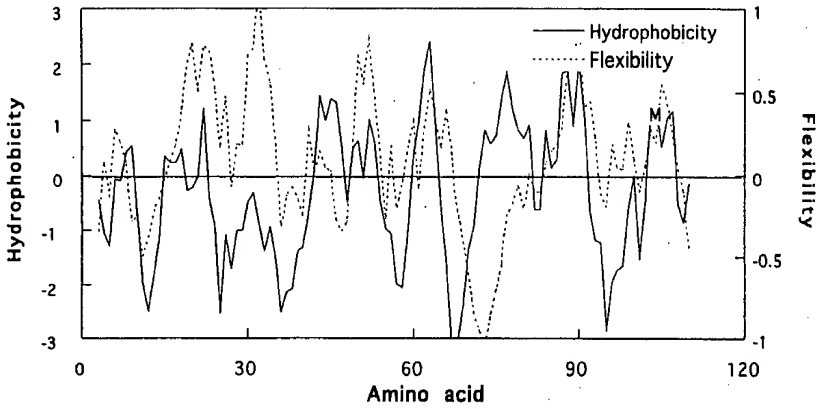
SEQUENCE

ALDTNYCFSSTEKNCCVRQLYIDFRKDLGWKWIHEPKGYHANFCLGPCPYIWSLDT-
QYSKVLALYNQHNPGASAAPCCVPQALEPLPIVYYVGRKPKVEQLSNMIVRSCKCS

REACTIVE SITES

.N.(6)			.D.(4)		.M.(1)	.Q.(5)	
5 TNY	66 YNQ	3 LDT	55 LDT	104 NMI	19 RQL	81 PQA
14 KNC	69 HNP	23 IDF			57 TQY	100 EQL
42 ANF	103 SNM	27 KDL			67 NQH	

HYDROFLEX PLOT



PREDICTED REACTIVITY AND DEGRADATION OF TGF-β

Inspection of the primary amino acid sequence reveals that TGF-β does not have any of the traditional hot spots for hydrolytic reactivity at neutral pH, in that Asn-Gly, Asn-Ser, Asp-Gly, and Asp-Pro are absent. TGF-β does have a single Met, and this is found in a region of predicted high hydrophobicity and decreased flexibility, possibly rendering this Met only weakly susceptible to oxidation. Recombinant TGF-β was remarkably stable and did not undergo noticeable chemical degradation in 0.1–1.0 mg/ml liquid formulations at pH 5 when stored for at least 1 year at 2–8°C. This is in good agreement with its hydroflex plot analysis, in that there are no traditional sites of reaction (except for a single Met in a nonflexible, hydrophobic environment).

Thrombopoietin (TPO) (332 residues)

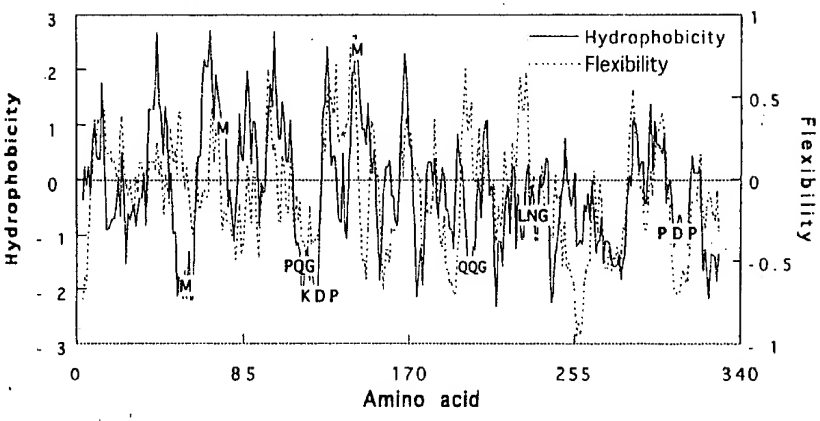
SEQUENCE

SPAPPACDLRVLSKLLRDSHVLHSRLSQCPEVHPLPTPVLLPAVDFSLGEWKTQMEE-
TKAQDILGAVTLLEGVMAARGQLGPTCLSSLLGQLSGQVRLLLGALQSLLGTQLP-
PQGRTTAHKDPNAIFLSFQHLLRGKVRFLMLVGGSTLCVRRAPPTTAVPSRTSLVLTL-
NELPNRTSGLLETNFTASARTTGSGLLKWQQGFRAKIPGLLNQTSRSLDQIPGYLNRI-
HELLNGTRGLFPGPSRRTLGAPDISSGTSDTGSLPPNLQPGYSPSPTHPPTGQYTLFP-
LPPTLPTPVVQLHPLLPDPSAPTPTPTSPLLNTSYTHSQNLSQEG

REACTIVE SITES

.N.(10)	.D.(9)	.M.(3)	.Q.(19)	
125 PNA	8 CDL	55 QME	28 SQC	201 WQQ
172 LNE	18 RDS	75 VMA	54 TQM	202 QQG
176 PNR	45 VDF	143 LML	61 AQD	214 NQT
185 TNF	62 QDI		80 GQL	221 DQI
213 LNQ	123 KDP		92 GQL	268 LQP
227 LNR	220 LDQ		96 GQV	282 GQY
234 LNG	252 PDI		105 LQS	298 VQL
266 PNL	259 SDT		111 TQL	326 SQN
319 LNT	305 PDP		115 PQG	330 SQE
327 QNL			132 FQH	

HYDROFLEX PLOT



PREDICTED REACTIVITY AND DEGRADATION OF THROMBOPOIETIN

There are several forms of TPO, including the natural full length sequence shown above (produced in either *E. coli* or CHO cells), as well as a number or truncated forms, some of which have also been pegylated. A preliminary stability analysis has been carried out on the full length "natural" molecule under physiological conditions (pH 7.4) (Lim *et al.*, 1996). Inspection of the primary amino acid sequence for TPO shows that the most reactive is predicted to be Asn-234 within the -LNG- motif, Asp-123 within the -KDP- motif, Asn-305 within the -PDP- motif, and Gln-115 within the -PQG- motif. The first of these, Asn-234, showed N-linked glycosylation in the CHO-derived molecule studied, and so this site was unavailable for reaction. All reside in a region predicted to be hydrophilic and flexible. The chemical stability of TPO was monitored by SEC and tryptic mapping. TPO deamidated at Asn-227 (in the -LNR- motif) and formed iso-Asp at Asp-220 (in the -LDQ- motif). Reaction at these sites did not alter TPO activity, nor did diketopiperazine formation at Ala-3 (des-Ser-Pro). Further, it was found that TPO aggregates, as well as oxidized TPO (using hydrogen peroxide), had little or no biological activity. The time required to achieve 90% rhTPO monomer (t_{90} shelf life) was determined to be greater than 2 years at 2–8°C.

Tissue Plasminogen Activator (human) (t-PA) (527 residues)

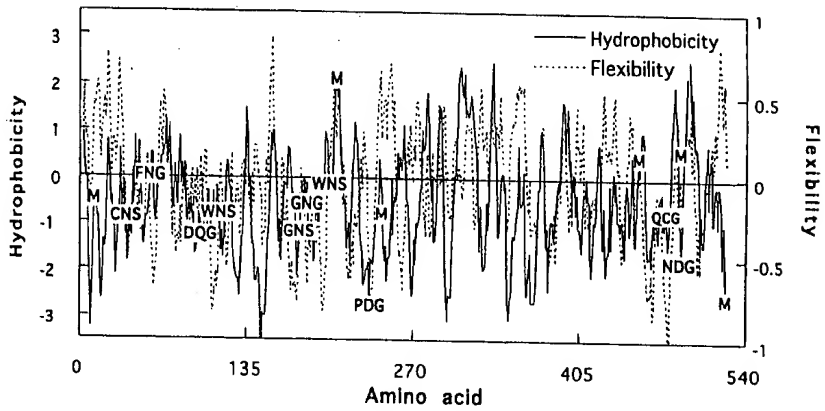
SEQUENCE

SYQVICRDEKTQMIYQQHQSWLRPVLRSNRVEYCWCNSGRAQCHSVPVKSCSEPR-
CFNGGTCQQALYFSDFVCQCPEGFAGKCCEIDTRATCYEDQGISYRGTWSTAESGAE-
CTNWNSSALAQKPYSGRRPDAIRLGLGNHNYCRNPDRDSKPWCYVFKAGKYSSE-
FCSTPACSEGNSDCYFGNGSAYRGTHSLTESGASCLPWNSMILIGKVYTAQNPSAQ-
ALGLGKHNYCRNPDGDAKPWCHMLKNRRLTWEYCDVPSCSTCGLRQYSQPQFR-
IKGGLFADIASHPWQAAIFAKHRRSPGERFLCGGILISSCWILSAAHCFQERFPPHHLT-
VILGRTYRVVPGEEEQKFEVEKYIVHKEFDDDTYDNDNALLQLKSDSSRCAQESS-
VVRTVCLPPADLQLPDWTECELSGYGKHEALSPFYSERLKEAHVRLYPSSRCTSQH-
LLNRTVTDNMLCAGDTRSGGPQANLHDACQGDSGGPLVCLNDGRMTLVGIISWGL-
GCGQKDVPGVYTKVTNYLDWIRDNMRP

REACTIVE SITES

.N.(23)		.D.(28)		.M.(6)	.Q.(26)	
29 SNR	248 KNR	8 RDE	366 DDT	13 QMI	3 YQV	271 SQP
37 CNS	370 DND	70 SDF	369 YDN	207 SMI	12 TQM	273 PQF
58 FNG	372 DNA	87 IDT	371 NDN	245 HML	16 YQQ	290 WQA
115 TNW	448 LNR	95 EDQ	380 SDS	455 NML	17 QQH	325 FQE
117 WNS	454 DNM	132 PDA	400 ADL	490 RMT	19 HQS	350 EQK
140 GNH	469 ANL	148 PDR	405 PDW	525 NMR	42 AQC	376 LQL
142 HNY	486 LND	150 RDS	453 TDN		63 CQQ	386 AQE
146 RNP	516 TNY	179 SDC	460 GDT		64 QQA	402 LQL
177 GNS	524 DNM	236 PDG	472 HDA		74 CQC	444 SQH
184 GNG		238 GDA	477 GDS		96 DQG	467 PQA
205 WNS		257 CDV	487 NDG		123 AQK	475 CQG
218 QNP		283 ADI	506 KDV		217 AQN	504 GQK
230 HNY		364 FDD	519 LDW		222 AQA	
234 RNP		365 DDD	523 RDN		268 RQY	

HYDROFLEX PLOT



PREDICTED REACTIVITY AND DEGRADATION OF t-PA

This serine protease is predicted to have several sites of possible degradation, notably deamidation at Asn-58 in the sequence -FNG, Asn-184 in the -GNG- sequence, Asn-177 in the sequence -GNS-, Asn-37 in the -CNS- motif, and Asn-117 in the -WNS- motif. Reaction was observed at most of these sites, including iso-Asp formation via deamidation of Asn-37 in the sequence -CNS- (Paranandi *et al.*, 1994). The Asn-Gly motifs are predicted to be reactive at neutral pH; the Asn-Ser motifs are also predicted to be reactive at 37°C based on synthetic peptide studies. When incubated at pH 7.3, 37°C, human recombinant t-PA accumulated 0.77 mol of iso-Asp per mol of t-PA over a 14-day period. All three sites appeared to be on the surface of the protein, and all three occurred in regions of the protein predicted to have higher than average chain mobility. It is interesting to note that Asn-184 within the -GNGS- motif was not susceptible to deamidation. The reason for this is straightforward; this Asn is glycosylated in t-PA expressed in CHO cells and so is unavailable for reaction. This protein is also glycosylated at Asn-117, possibly accounting for its lack of reaction at this motif (although the -WNS- is not necessarily reactive). Although this molecule has numerous Mets, no reports of Met oxidation were reported. These hydrolysis reactions did not limit the shelf life in that this molecule is subject to another, more rapid, degradation pathway (Nguyen and Ward, 1993b). This molecule is a serine protease and so is subject to autocatalytic degradation; because of this, it is formulated as a lyophilized powder and reconstituted before use.

Trypsin (bovine) (223 residues)

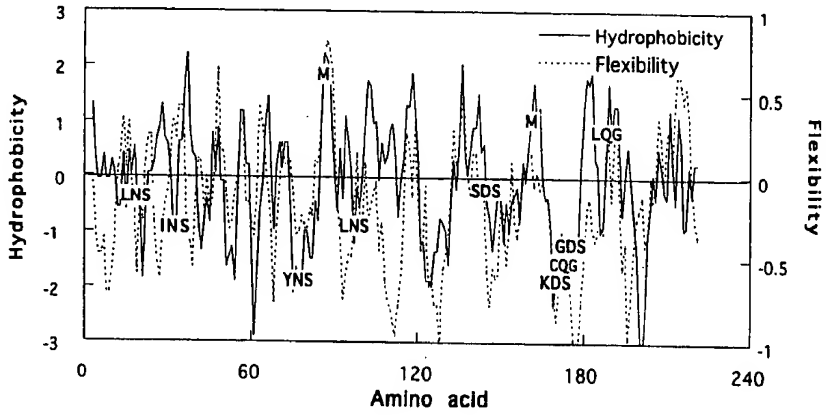
SEQUENCE

IVGGYTCGANTVPYQVSLNSGYHFCGGSLINSQWVVSAAHCYKSGIQVRLGEDN-
INVVEGNEQFISASKSIVHPSYNSNTLNNDIMLIKLKSAASLNSRVASISLPTSCASAG-
TQCLISGWGNTKSSGTSYPDVLKCLKAPILSDSSCKSAYPGQITSNMFCAGYLEGG-
KDSCQGDSGGPVVCSGKLQGIVSWGSGCAQKNKPGVYTKVCNYVSWIKQTIASN

REACTIVE SITES

.N.(16)		.D.(6)	.M.(2)	.Q.(10)	
10 ANT	82 LNN	53 EDN	86 IML	15 YQV	199 AQK
19 LNS	83 NND	84 NDI	160 NMF	33 SQW	218 KQT
31 INS	97 LNS	133 PDV		47 IQV	
54 DNI	123 GNT	145 SDS		63 EQF	
56 INV	159 SNM	171 KDS		115 TQC	
61 GNE	201 KNK	176 GDS		155 GQI	
77 YNS	211 CNY			174 CQG	
79 SNT	223 SN			188 LQG	

HYDROFLEX PLOT



PREDICTED REACTIVITY AND DEGRADATION OF TRYPSIN

Inspection of the amino acid sequence of trypsin shows that there are several Asn-Ser motifs, all located in hydrophilic, flexible regions. Trypsin also contains two Met residues, both located in hydrophobic, inflexible regions. Based on this, it is likely that trypsin may show deamidation or cyclic imide formation at any (or all) of the -XNS- motifs. Interestingly, Gln-174 is found in a region that is fairly flexible and hydrophilic, although this motif contains a cysteine that, when forming a disulfide bridge, may reduce the local flexibility dramatically, rendering it fairly unreactive. An elegant NMR study showed that three residues were prone to microheterogeneity (in the form of a deamidated product): Asn -31 (-INS), Asn-77 (-YNS-), and Asn-97 (-LNS) (revised numbering system to make the N-termini start at 1) (Kossiakoff, 1988). None of the other 13 Asn residues showed reactivity under the experimental conditions used. Of note, Asn-19 (-LNS-) did not show microheterogeneity, even though it has the same motif as Asn-97. This is another clear-cut demonstration that conformational aspects are crucial for deamidation in proteins. In this study, it was not determined if deamidation occurred prior to crystallization or if it occurred during the 1-year period of crystal growth and data

collection. Nevertheless, these data show that trypsin does not degrade at sites other than the predicted hot spots.

Vascular Endothelial Growth Factor (VEGF) (165 residues)

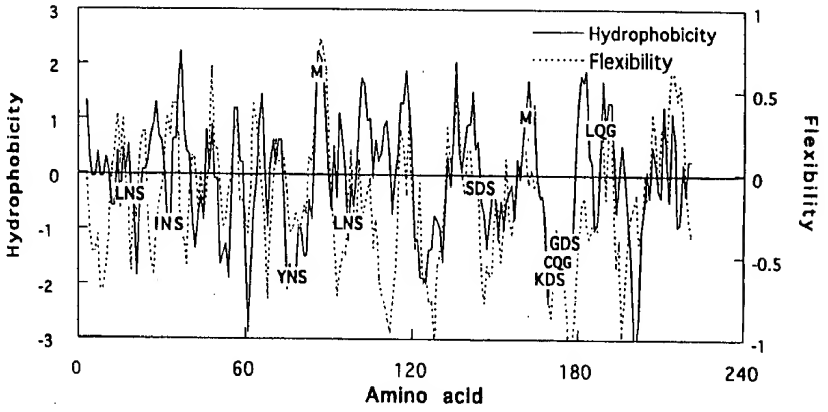
SEQUENCE

APMAEGGGQNHHEVVKFMDVYQRSYCHPIETLVDIFQEYPDEIEYIFKPSCVPLMR-
CGGCCNDEGLECVPTEESNITMQIMRIKPHQGQHIGEMSFLQHNKCECRPKKDRA-
RQENPCGPCSERRKHLFVQDPQTCKCSCKNTDSRCKARQLELNERTCRCDKPRR

REACTIVE SITES

.N.(7)	.D.(8)	.M.(6)	.Q.(11)
10 QNH	19 MDV	3 PMA	9 GQN
62 CND	34 VDI	18 FMD	22 YQR
75 SNI	41 PDE	55 LMR	37 FQE
100 HNK	63 NDE	78 TMQ	79 MQI
115 ENP	109 KDR	81 IMR	87 HQG
141 KNT	131 QDP	94 EMS	89 GQH
154 LNE	143 TDS		98 LQH
	161 CDK		113 RQE
			130 VQD
			133 PQT
			150 RQL

HYDROFLEX PLOT



PREDICTED REACTIVITY AND DEGRADATION OF VEGF

Inspection of the amino acid sequence of VEGF shows that there are few hot-spot motifs, and the two that exist (Gln-Gly and Asp-Pro) are not predicted to be as reactive as Asn-Gly or Asn-Ser. VEGF has a Pro at position 2 (APM ...), suggesting that this molecule might undergo diketopiperazine formation. It also has several Met residues that may oxidize. The degradation of VEGF in aqueous solutions from pH 5 to 7 has been determined (Keyt and Cleland, 1995). From pH 5 to 6, the major degradation route at accelerated conditions of 40°C was deamidation at Asn-10 in the -QNH- motif to give a variety of products, as yet to be determined. At higher pH, proteolysis and additional deamidation were observed but not fully characterized. At or above pH 6.5, some diketopiperazine formation was observed under accelerated conditions of 40°C for 4 weeks, giving the expected reaction product, des-Ala-Pro VEGF.

4. STATISTICAL ANALYSIS OF PROTEIN DEGRADATION SITES IN AQUEOUS SOLUTION

These data show that the primary reaction of proteins at pH 4.5–7.5 occurs largely at Asn and Asp within these motifs: -Asn-Gly-, -Asn-Ser-, -Asp-Gly-, and to a lesser extent -Gln-Gly-, -Asp-Pro-, and -Met-. A few proteins, however, react at sites other than these and are exceptions to the rule. These proteins react at sites that are deemed "unreactive" sites (such as at Asn 52 in the -LND- motif in CD4), based largely upon data obtained in small model peptides. There are several reasons why proteins may show unusually high reactivity at these non-hot-spot sites:

(i) The motif is in the "correct" conformation for reaction to occur.
(ii) Reaction may be due to enzymatic catalysis (traces of unwanted proteases in the purified protein product).
(iii) The protein degrades under the harsh conditions of isolation and work-up.
(iv) The DNA encodes for both the parent and the product forms (such as encoding for Asn and Asp in cholera B subunit, depending on the strain studied), so both isoforms are expressed.

The first reason that the reactive motif is held conformationally in a geometry that facilitates reaction is an often-touted explanation for "non-hot-spot" protein reactivity. But is it the only reason? For example, some proteins deamidate faster under work-up conditions than in pH 7.4 buffer at 37°C, suggesting that enzyme catalysis [point (ii)] is operational. Enzymatic catalysis may not only cause an acceleration in the rate of deamidation, but may also promote deamidation at motifs that would have otherwise been unreactive at pH 7.4 and 37°C. There are also numerous examples in the literature (see below) of non-hot-spot protein degradation coming from publications where the work-up of the protein was carried out under fairly harsh conditions and often without controls. Finally, a few concrete examples

of protein microheterogeneity due to different DNA coding have been reported and provide an elegant rationale for apparent non-hot-spot protein degradation.

To sort through this, we have constructed a table summarizing protein reactivity in aqueous solution, based on whether or not the primary degradation pathway was observed to be at one or more of the predicted hot spots (Table II). This table contains information on the degradation behavior of 73 proteins, including information on the frequency of hot spots of each type. Most of these proteins show degradation, and this degradation information is available as to whether it has been observed to occur at a hot spot, at some other motif, or through oxidation. Information on 54 of the 73 proteins was obtained under formulation conditions (i.e., carefully controlled formulation-type studies where the reaction catalyst and the degradation kinetics are fairly well understood and not complicated by the initial protein quality or enzymatic degradation); for 21 proteins, degradation information pertains to behavior determined under "work-up" conditions (where degradation may also occur due to the work-up process, enzymatic catalysis in a biological milieu, or to fermentation degradation before isolation). Two proteins, calmodulin and interleukin 2, have been studied under both formulation and work-up conditions and so are included in both data sets. An additional column was added to address oxidation as the primary degradation pathway (under either formulation or work-up conditions), and no distinction between formulation and work-up was made as few proteins undergo oxidation as the primary degradation pathway in aqueous solution. A final column summarizes those proteins that react as predicted, based on the hypothesis that the primary degradation pathway occurs at one of the predicted hot spots for reaction.

Inspection of the data in Table II shows that there are several proteins that react at non-hot-spots; these are traditionally thought to be exceptions to the rule of protein reactivity. For example, of the 54 proteins that were studied under formulation conditions, 32 (~60%) showed primary reaction at hot-spot sites and 22 (~40%) showed reaction at non-hot-spot sites. Of the 21 proteins that were studied under work-up conditions, 10 (~48%) showed primary reaction at hot-spot sites and 11 (~52%) show reaction at non-hot-spot sites. Closer scrutiny of the data, however, shows that some of these differences arise because many proteins are devoid (or have very few) of the traditional hot spots, so when degradation is observed it is ultimately at a non-hot-spot site of degradation. To account for this, a column in the table called "predicted reactivity" was added. An absence of an X in the last column indicates proteins that are the truly unusual cases of protein degradation (i.e., proteins that degrade at non-hot-spots, even though there are traditional hot spots which remain unreactive). Inspection of the table shows that there is a slightly higher tendency to observe non-hot-spot protein degradation when studied under work-up conditions, although this may not be statistically valid (i.e., $p < 0.05$) because of the limited data subset size. In several of these cases, these exceptions are found under conditions where work-up reaction may occur, or where harsh conditions of protein isolation may account for some of the observed degradation. From a formulator's point of view, reaction of proteins at 37°C at pH 7.4 in a biological milieu containing enzymes

Table II. Statistical Analysis of Protein Degradation Pathways[a]

Protein	# NG	# NS	# DG	# DP	# QG	# M	Hot-spot formul.	Other formul	Hot-spot work-up	Other work-up	Oxidation	Pred reac
Adrenocorticotropin	1					1	X					X
Agglutinin	1		1		2	2			X			X
Aldolase	2	1	4		2	3				X		
Amylin antagonist								X				X
Amyloid-related serum protein		1	1	1		2				X		
Angiogenin	2	1			1	1	X					X
Anti-HER-2 heavy chain	2	3	3	1	1	4				X		X
Anti-HER-2 light chain		1			2	1						X
Antibody 4D5 heavy chain	2	3	3	1	2	5	X					X
Antibody 4D5 light chain		1			2	1	X					X
Antibody 17-1A heavy chain	1	3	2	1	2	7		X				
Antibody 17-1A light chain	1	2	1		1	4	X					X
Antibody light chain kappa	1	3	1		1	4	X					X
Antibody OKT3 heavy chain	1	2	1			5	X					X
Antibody OKT3 light chain	1	3	1	1	2	9	X					X
Antibody OKT4 heavy chain	2	3	2	2	1	5	X					X
Antibody OKT4 light chain		1			2	1		stable				X
Atrial natriuretic peptide		1				1	X				X	X
Brain derived neurotrophic factor		1		1		3	X	X				X
Calbindin	1		1	1			X					X
Calmodulin	2		6			9	X		X			X
Carbonic anhydrase	3		3	2	1	1			X			X
CD4		2		1	4	4		X				
CD4-PE40	1	2	1	2	5	1					X	X
Chloroperoxidase		2	1	1	1	3				X		
Cholera toxin B subunit	1					3			X			X
Ciliary neurotrophic factor			3		1	4		X				
Crystallin-A			1	1		3				X		
Cytochrome *c*						2		X				X
DNase		1	1			5	X					X
Epidermal GF (human)		1	2			1	X				X	X
Epidermal GF (murine)	1	1	2			1	X					X
Erythrocyte protein 4.1	1		2	2	1	6			X			X
Fibroblast GF acidic	2		3			2		X				
Fibroblast GF basic	1		4	2		2	X					X
Glucagon					1	1				X		X
Granulocyte CSF			1		3	3		X				
Growth hormone (bovine)		1	1			5	X					X
Growth hormone (human)		2				3	X					X
Growth hormone (porcine)		1	1			3	X					X
Growth hormone rel factor		1			1		X					X
Hemoglobin				1		2				X		
Hirudin			2		1				X			X
Histone					2	1				X		
HXGT		1				6				X		
Insulin								X				X
Insulin-like growth factor I						1		X			X	X
Insulinotropin											X	
Interferon-α-2b						5					X	X
Interferon-β	1					3				X		X
Interferon-γ	1	1		1		4	X				X	X
Interleukin 1-RA					1	4		X			X	X

(*continued*)

Table II. (*Continued*)

Protein	NG #	NS #	DG #	DP #	QG #	M #	Hot-spot formul.	Other formul.	Hot-spot work-up	Other work-up	Oxidation	Pred reac
Interleukin 1α						3		X				X
Interleukin 1β (human)				1	2	6		X				X
Interleukin 1β (murine)	1	1	1	2	1	4	X					X
Interleukin 2	1					4	X	X		X		X
Interleukin 11			1	2		2		X			X	X
Lung surfactant						1						
Lysozyme	1	1	3			2	X				X	X
Myelin basic protein			1		3	2			X			X
Neocarzinostatin		1	2	1	1			X				
Nerve growth factor		2	1	1		2	X				X	X
Parathyroid hormone		1				2				X		
Relaxin						2		X				X
RNAase A	1					4			X			X
RNAase U2	2		1			1			X			X
Secretin			1			1	X					X
SHMT	2	1	2	2	1	8	X					X
Tissue factor-243		1				1		stable				X
Tissue growth factor-β						1		stable				X
Tissue plasminogen activator	2	4	2		2	6	X					X
Trypsin		4			2	2	X					X
VEGF				1	1	6		X				X

[a]Note: Data for antibody E25 and TPO were not included in this analysis. It is not expected that the omission of these proteins changes the results significantly.

that may cause protein degradation does not mimic optimal formulation reaction conditions.

It has long been known that many proteins react at the traditional hot spots, but the predictive value of this general knowledge has not yet been tested. For example, what is the probability of primary reaction at Asn-Ser for a new, unstudied protein? The main objective of our analysis is to determine if the frequency of motifs of a particular type affects the propensity for proteins to degrade at a hot spot or at some other motif. In addition, the association between oxidation and the frequency of methionine residues was investigated.

In assessing hot-spot degradation, the following evaluations were conducted separately for proteins studied under formulation conditions and those studied under work-up reaction conditions. Note that the number of proteins used in the statistical analysis differs slightly from the total number of proteins in this compendium, largely because some proteins are structurally similar and so would unduly weight the analysis if all were used. For each hot-spot type, the frequency distribution was compared between proteins which exhibited degradation at a hot spot and those which did not. This comparison was carried out by formal contingency table analysis. A two-tailed Fisher's exact test was used to assess significance of the association between degradation at a hot spot and frequency of a particular motif for each of the five hydrolytic hot-spot types. An additional evaluation investigated this association with the frequency of hot spots of any type. The results are shown in Appendix A. For proteins studied under formulation conditions, results are summarized in Tables AIa–f and Figs. A1a–f, respectively. The corresponding Tables AIIa–f and Figs. AIIa–f present results for the proteins investigated under work-up conditions. The relationship between oxidation and frequency of Met residues is summarized in Table AIII; Fig. A3 provides a graphical summary.

By convention, p-values less than 0.05 are reported as representing a statistically significant association; however, all p-values should be interpreted with caution, since not all 73 of the proteins analyzed can be considered to provide independent information (e.g., results for the antibodies are unlikely to be independent because of sequence similarity outside of the CDR region). Since only 21 proteins were studied under work-up conditions, p-values were not reported for the association between degradation and hot-spot frequency; small sample sizes and discreteness of the distributions involved render formal hypothesis tests suspect in this case. For proteins studied under work-up conditions, graphs and tables are provided for descriptive purposes only.

The frequency distribution of Met residues was compared across proteins which did and did not undergo oxidation in similar fashion. This analysis was carried out for all 73 proteins combined, without distinguishing between those studied under formulation and work-up conditions. The analyses support the following conclusions:

- There is a pronounced shift in the frequency distribution of the -Asn-Gly- motif among proteins which degrade at a hot spot under formulation condi-

tions. The majority (86%) of proteins not exhibiting hot-spot degradation lack an -Asn-Gly- motif, whereas a majority (59%) of those which do degrade at a hotspot have at least one -Asn-Gly- motif (Table AIa/Fig. AIa). The p-value of 0.004 suggests that this is not a random event.

- Similarly, the frequency of the -Asn-Ser- motif appears positively associated with hot-spot degradation under formulation conditions (Table AIb/Fig. A1b). The majority (71%) of proteins not exhibiting hot-spot degradation lack an -Asn-Ser- motif, whereas nearly half (40%) of those which do degrade at a hot spot have at least one -Asn-Ser- motif. The p-value for this analysis is 0.002 suggesting that this is not a random event.
- The tendency to degrade at a hot spot under work-up conditions does not appear to be associated with frequency of the other motifs: -Asp-Gly- (p = 0.22), -Asp-Pro- (p = 1.00), or -Gln-Gly- (p = 0.27. (Tables AIc–e/Figs. AIc–e). There may be several reasons for this. First, if these motives are unreactive on the time scale studied, then large values of p will be obtained. Second, if reaction goes undetected because of experimental difficulty (such as might be the case for iso-Asp formation from Asp), this will also result in an apparent lack of association.
- Not surprisingly, there is a significant association between degradation at a hydrolytic hot spot and the overall number of hot spots. The significance of this association, however, may be driven in part by the structural zero in Table AIf (proteins without any hot spot cannot degrade at a hot spot).
- In general, patterns for degradation under work-up conditions appear similar to those for the proteins studied under formulation conditions. An exception may be the pattern of -Asn-Ser- motifs (compare Tables AIIb and AIb), although the sample size (20) is too small to draw definitive conclusions.
- Since oxidation occurred for relatively few (11) of the 73 proteins, the ability to assess the relationship to the frequency of Met residues is limited. The p-value of 0.62 shows that there is no correlation between oxidation and the presence of Met (i.e., many proteins containing Met do not oxidize).

5. GENERAL CONCLUSIONS REGARDING PROTEIN DEGRADATION IN AQUEOUS SOLUTION

This literature compilation on the chemical reaction of proteins was assembled to establish boundaries to the reactivity of Asn, Asp, Gln, and possibly Met, in the context of neighboring amino acid sequence, regional hydrophobicity, and backbone flexibility. An extensive review of the literature, as well as several unpublished reports, afforded numerous proteins that selectively hydrolyze, deamidate, undergo iso-Asp formation, or oxidize in aqueous solution. Inspection of the primary amino acid sequence alone gives a modest indicator of the most reactive motifs; it was found

that the general rules already established predict the majority of reactive sites observed (Asn-Gly, Asn-Ser, Asp-Gly, Gln-Gly, Asp-Pro, and Met). Of the proteins of which we have compiled reliable degradation data, only 5 (CD4, CNTF, acidic-FGF, GCSF, and neocarzinostatin) degraded primarily at "unusual" sites of degradation, and not at the available and predicted hot spots. We calculated the hydropathy/flexibility plots (termed "hydroflex" plots) to provide a way of further examining the degradation of peptides and proteins. By doing so we found that some residues predicted to react based on their amino acid sequence (but did not react experimentally) were calculated to be in hydrophobic regions of limited flexibility. Further, the hydrolytic protein degradation studies have been carried out under two types of conditions: those carefully controlled studies in aqueous solution at near-neutral pH where the integrity of the initial protein was well known (termed formulation studies) and those where degradation was observed after isolation from biological media or where the protein may have degraded upon work-up (herein termed, work-up studies). We present several examples of work-up degradation that do not adhere to the above rules (based on the predicted hot spots), possibly because of enzymatic catalysis or extreme reaction conditions used for protein isolation and purification. For these reasons, work-up degradation results should not be used to predict protein reactivity in aqueous formulations. Finally, the prediction of Met reactivity based on primary amino acid sequence was not successful, possibly because of the limited protein oxidation data available, because of the complex nature of protein oxidation by a variety of different oxidative catalysts (Knepp *et al.*, 1996), or because protein conformation prohibits reaction of Mets found in the protein core. Some general conclusions are emphasized.

1. Data used to predict protein reactivity in aqueous formulations should be carefully scrutinized before making general conclusions. Several of the exceptions to the rule for protein degradation in aqueous solution come from examples in the literature where the nature of the "unusual" degradation is unknown and may be caused by enzymatic degradation, heterogeneity of protein expression at the gene level, or hydrolytic or oxidative degradation upon work-up. These examples are not representative test cases for protein degradation in aqueous formulations.
2. Hydropathy seems to be a better predictor for protein degradation than does calculated flexibility. Inspection of more than 70 hydroflex plots shows that most of the reactive hot spots lie in regions predicted to be hydrophilic. In large part this is due to the nature of the calculation (for example, Asn, Asp, and Gln have large negative Kyte parameters, lowering the overall value of calculated hydrophobicity). For example, the literature average (over 500,000 protein entries included in this calculation) hydropathy for all residues is −0.32 (statistically corrected for the amount of each amino acid found in nature); similarly the hydropathy values for NG, NS, DG, QG, DP, and M (again statistically corrected and using a window of six amino acids in

the hydropathy calculation) are −0.81, −0.89, −0.84, −0.84, −1.06, and 0.04, respectively.

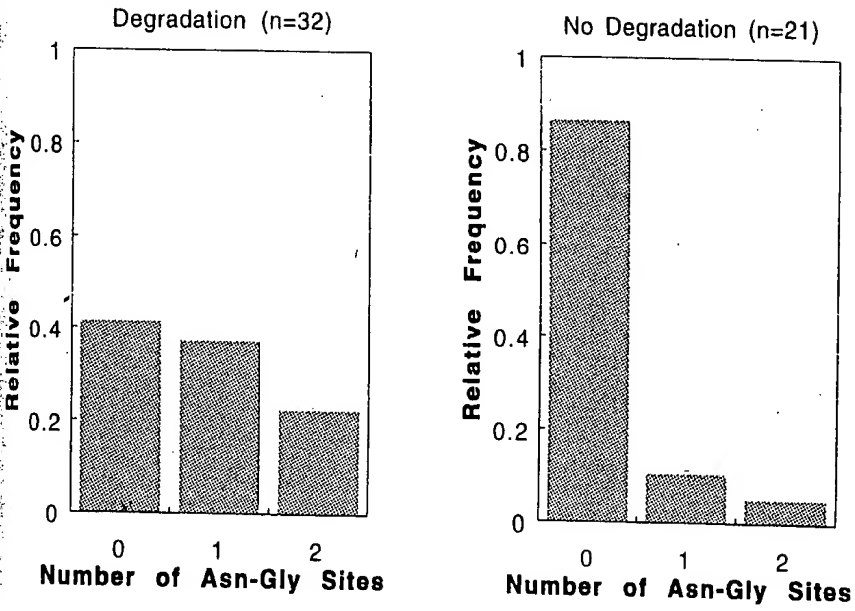
3. Met oxidation is not a major pathway for degradation for most proteins and is difficult, if not impossible, to predict based on primary sequence alone. Many of the proteins studied had several Met amino acids and yet did not show oxidative degradation; other proteins, however, showed oxidation as the primary degradation pathway. A few proteins also showed oxidation as a minor degradation pathway. For the handful of proteins that showed oxidation at Met, there was no observable correlation between Met reactivity and hydrophobicity. Again, the average Met hydrophobicity (statistically corrected) over the entire database was 0.04; the value, along with the standard deviation of the calculated hydropathy for reactive Mets was 0.2 ± 1.6.
4. There appears to be a fairly good correlation between degradation at hot spots and the number of available hot spots for reaction. Proteins with only a few hot spots (or the lesser reactive hot spots such as Gln-Gly) tended to react at non-hot-spot sites, whereas proteins with numerous hotspots, particularly if they included Asn-Gly, Asn-Ser, and Asp-Gly, tended to react primarily at these sites. There were exceptions to this conclusion, but they were few and did not represent the norm.
5. Based on the above, the design of protein stability experiments in aqueous formulations should focus initially on the identity of hot-spot degradation pathways, with emphasis on Asn-Gly and Asn-Ser (as applicable). Less attention should be focused on Gln-Gly, as this motif appears to be less reactive than Asp-Gly, Asp-Pro, or even Met.

APPENDIX A

There was a pronounced shift in the frequency distribution of the -Asn-Gly- motif among proteins which degraded at a hot spot under formulation conditions. The majority (86%) of proteins not exhibiting hot-spot degradation lacked an -Asn-Gly- motif, whereas a majority (59%) of those which degraded at a hot spot had at least one -Asn-Gly- motif. The p-value of 0.004 suggests that this is not a random event.

Table AIa. Hot-Spot Degradation under Formulation Conditions by Frequency of Asn-Gly motifs

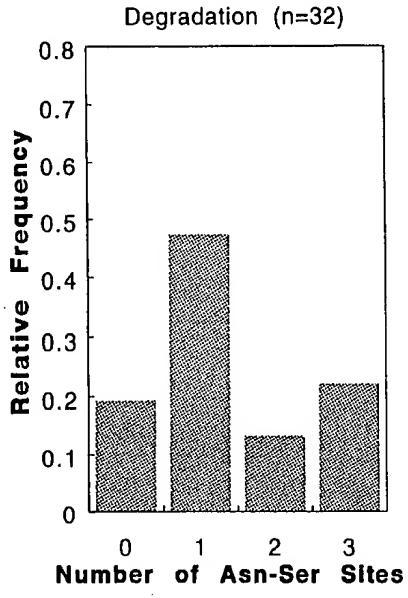
	Frequency					
Degrades	0	1	2	3+	Total	
No	18 (86%)	2 (10%)	1 (5%)		21	
Yes	13 (41%)	12 (37%)	7 (22%)		32	
Total	31 (59%)	14 (26%)	8 (15%)		53	$p = 0.004$

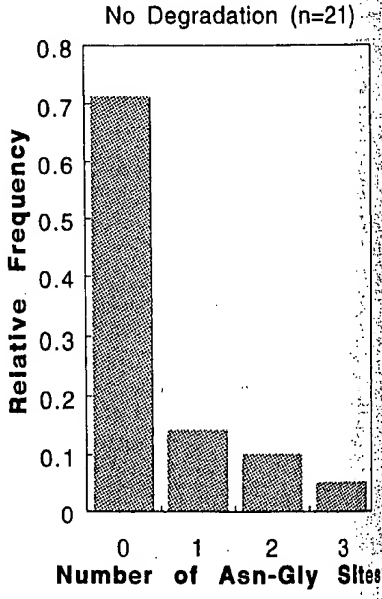


The frequency of the -Asn-Ser- motif appeared positively associated with hot-spot degradation under formulation conditions. The majority (71%) of proteins not exhibiting hot-spot degradation lacked an -Asn-Ser- motif, whereas nearly half (40%) of those which degraded at a hot spot had at least one -Asn-Ser- motif. The *p*-value for this analysis is 0.002, suggesting that this is not a random event.

Table AIb. Hot-Spot Degradation under Formulation Conditions by Frequency of Asn-Ser motifs

	Frequency					
Degrades	0	1	2	3+	Total	
No	15 (71%)	3 (14%)	2 (10%)	1 (5%)	21	
Yes	6 (19%)	15 (47%)	4 (13%)	7 (22%)	32	
Total	21 (40%)	18 (34%)	6 (11%)	8 (15%)	53	$p = 0.002$

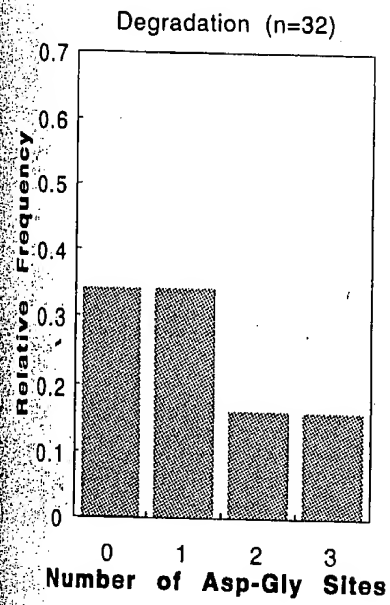


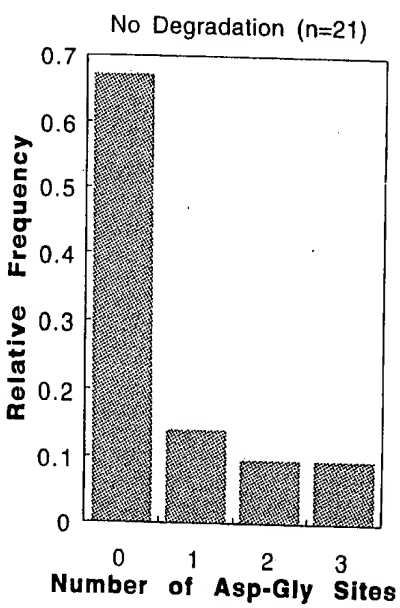


The tendency to degrade at the -Asp-Gly- hot spot under formulation conditions did not appear to be associated with frequency of the -Asp-Gly- motif. The *p*-value for this analysis was 0.22, suggesting that this was a random event. The reaction of Asp-Gly is often difficult to detect because of experimental difficulty (such as iso-Asp formation from Asp), and may account, at least in part, for the apparent lack of association.

Table AIc. Hot-Spot Degradation under Formulation Conditions by Frequency of Asn-Gly Motifs

	Frequency					
Degrades	0	1	2	3+	Total	
No	14 (67%)	3 (14%)	2 (9.5%)	2 (9.5%)	21	
Yes	11 (34%)	11 (34%)	5 (16%)	5 (16%)	32	
Total	25 (47%)	14 (26%)	7 (13%)	7 (13%)	53	$p = 0.22$

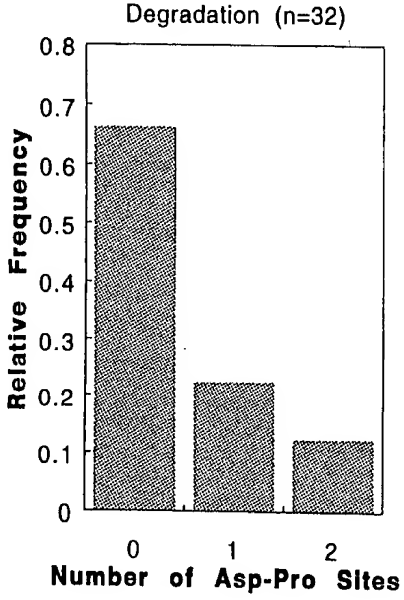


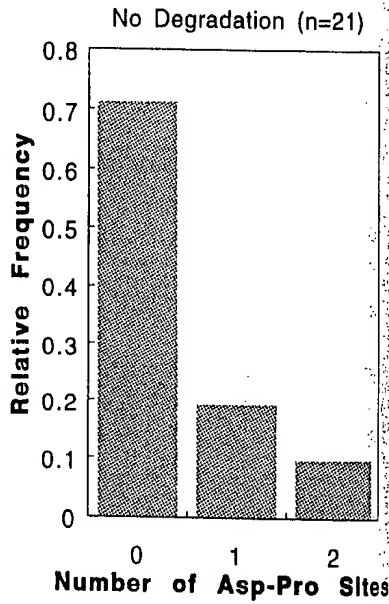


The tendency to degrade at the -Asp-Pro- hot spot under formulation conditions did not appear to be associated with frequency of the -Asp-Pro- motif. The *p*-value for this analysis was 1.00, suggesting that this was a random event. The reaction of Asp-Pro is favored at low pH's and becomes less favorable as the pH is raised. Because many formulations are made at near-neutral pH, it is likely that this reaction is minimized and may account, at least in part, for the apparent lack of association.

Table AId. Hot-Spot Degradation under Formulation Conditions by Frequency of Asp-Pro Motifs

	Frequency					
Degrades	0	1	2	3+	Total	
No	15 (71%)	4 (19%)	2 (9.5%)		21	
Yes	21 (66%)	7 (22%)	4 (12%)		32	
Total	36 (68%)	11 (21%)	6 (11%)		53	$p = 1.00$

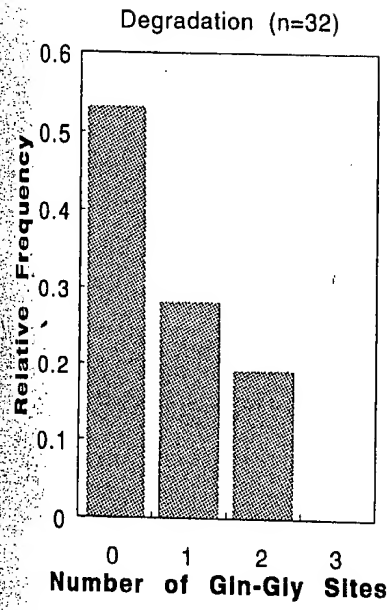


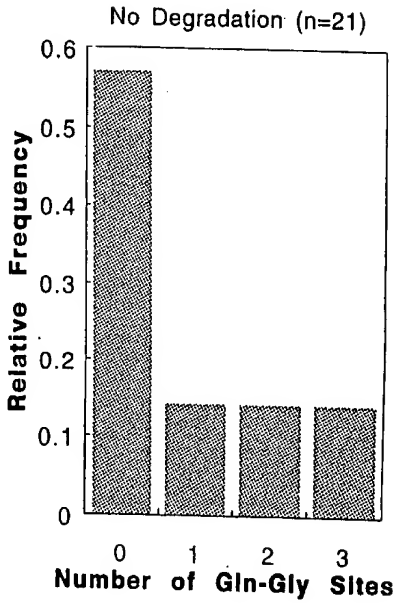


The tendency to degrade at the -Gln-Gly- hot spot under formulation conditions did not appear to be associated with frequency of the -Gln-Gly- motif. The *p*-value for this analysis was 0.27, suggesting that this was a random event. The reaction of Gln-Gly was observed only rarely in all of the proteins tabulated, even though this motif was found often. It is likely that the apparent lack of association of this motif with reactivity is caused by just that—lack of reactivity under formulation conditions.

Table AIe. Hot-Spot Degradation under Formulation Conditions by Frequency of Gln-Gly Motifs

	Frequency					
Degrades	0	1	2	3+	Total	
No	12 (57%)	3 (14%)	3 (14%)	3 (14%)	21	
Yes	17 (53%)	9 (28%)	6 (19%)	0	32	
Total	29 (55%)	12 (23%)	9 (17%)	3 (6%)	53	$p = 0.27$

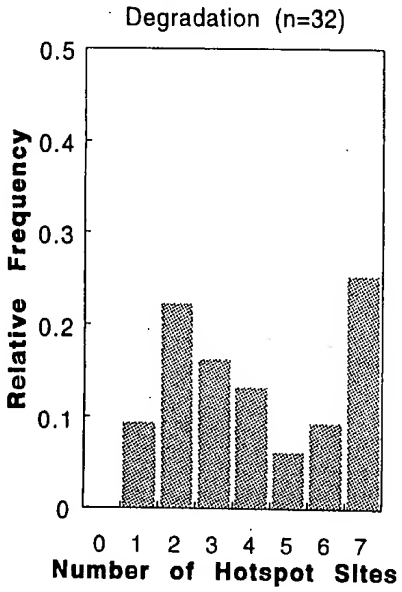




The tendency to degrade at any hot spot under formulation conditions was tightly associated with the frequency of hot spots. The p-value for this analysis is 0.0004 suggesting that this was not a random event. The small p-value for this association may in part be driven by the structural zero in the degradation plot; no degradation at a hot spot is possible if the protein does not have a hot spot.

Table AIf. Hot-Spot Degradation under Formulation Conditions by Frequency of Any Motif

	Frequency									
Degrades	0	1	2	3	4	5	6	7+	Total	
No	9	2	0	3	2	2	0	3	21	
Yes	0	3	7	5	4	2	3	8	32	
Total	9	5	7	8	6	4	3	11	53	p = .0004



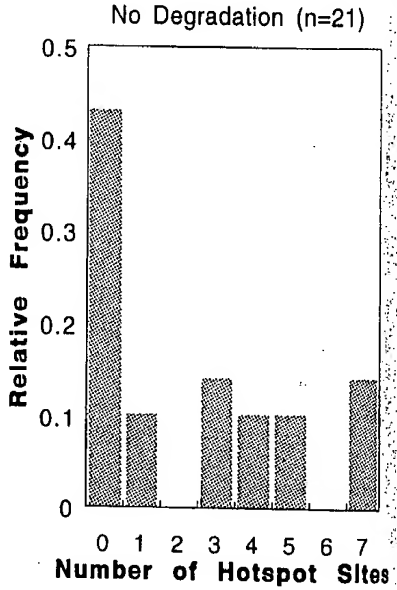
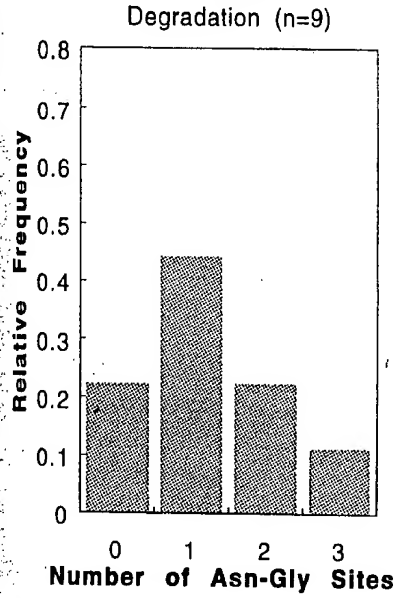


Table AIIa. Hot-Spot Degradation under Work-up Conditions by Frequency of Asn-Gly Motifs

	Frequency					
Degrades	0	1	2	3+	Total	
No	8	2	1	0	11	
Yes	2	4	2	1	9	
Total	10	6	3	1	20	p = 0.103



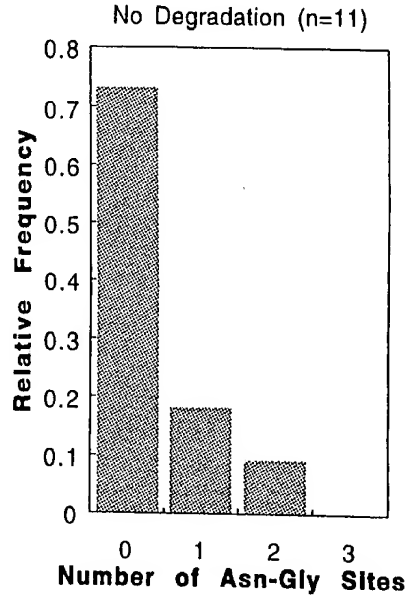
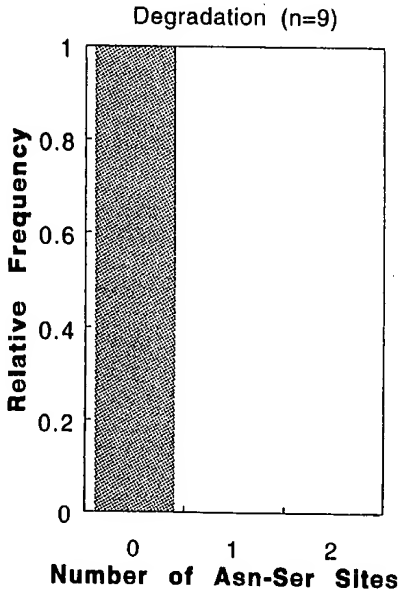


Table AIIb. Hot-Spot Degradation under Work-up Conditions by Frequency of Asn-Ser Motifs

	Frequency				
Degrades	0	1	2	3+	Total
No	6	4	1		11
Yes	9	0	0		9
Total	15	4	1		20

$p = 0.056$



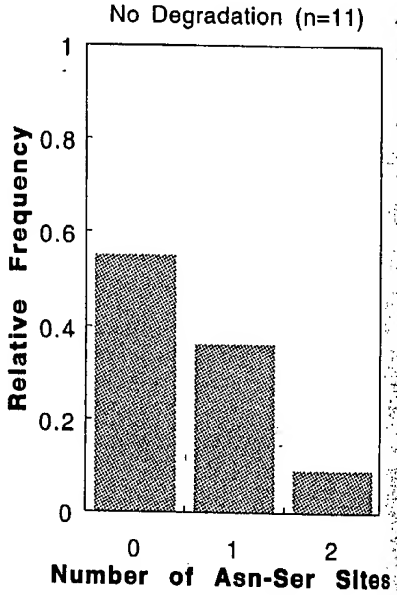
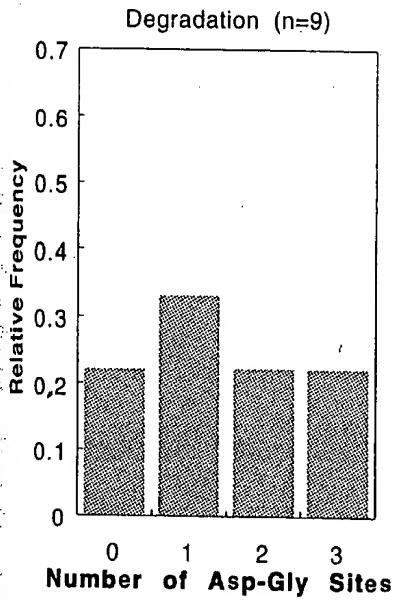


Table AIIc. Hot-Spot Degradation under Work-up Conditions by Frequency of Asp-Gly Motifs

	Frequency				
Degrades	0	1	2	3+	Total
No	7	3	0	1	11
Yes	2	3	2	2	9
Total	9	6	2	3	20

$p = 0.098$



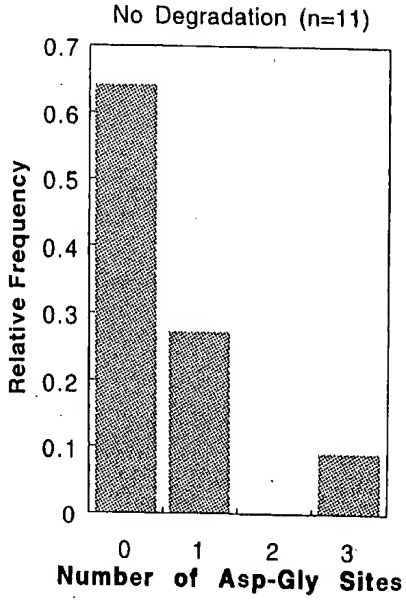
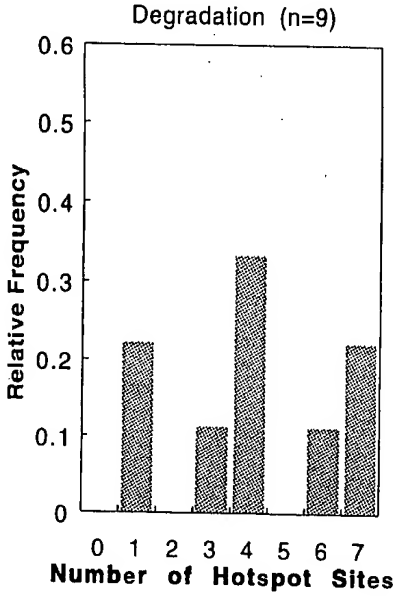
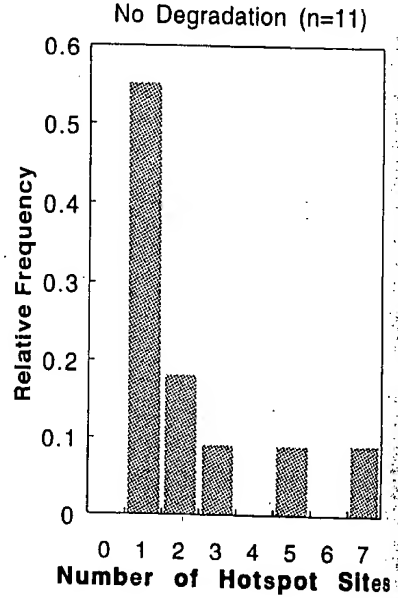


Table AIIf. Hot-Spot Degradation under Work-up Conditions by Frequency of Any Motif

Degrades	0	1	2	3	4	5	6	7+	Total	
No	0	6	2	1	0	1	0	1	11	
Yes	0	2	0	1	3	0	1	2	9	
Total	0	8	2	2	3	1	1	3	20	$p = 0.133$

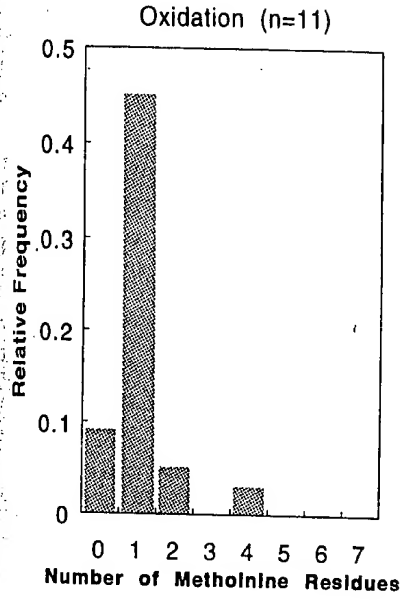


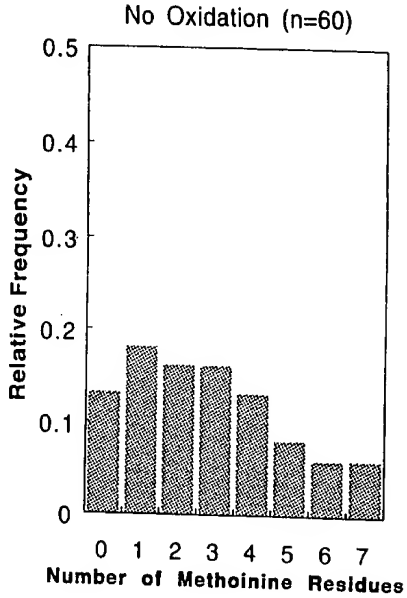


These plots are intriguing and provocative, in that the did not find a correlation of oxidation with number of Met residues. This is perhaps somewhat surprising, in that one might have intuitively expected that proteins with many Met residues maight be more prone to oxidation that those with few. Indeed, there existed a single example where oxidation was the predominant pathway, and yet the protein is devoid of Met (oxidation occurred at Trp).

Table AIII. Oxidation by Frequency of Met Residues

Oxidation?	0	1	2	3	4	5	6	7+	Total	
No	8	11	10	10	8	5	4	4	60	
Yes	1	5	3	0	2	0	0	0	11	
Total	9	16	13	10	10	5	4	4	71	$p = 0.062$





ACKNOWLEDGMENTS. This chapter would not have come together without the help of Jessica Burdman and Milianne Chin. Several others have contributed to this compendium: Sid Advant (*Protein Design Labs*), Dana Aswad (*UC Irvine*), Nancy Babour (*Bristol Myers*), John Battersby (*Genentech, Inc.*),Tom Bewley (*Genentech Inc.*), Ron Borchardt (*University of Kansas*), Bill Charman (*Monash University*), Diane Corbo (*R. W. Johnson*), Bim Dhingra (*Circa Pharmaceuticals Inc.*), Marcia Federici (*SKB*), John Frenz (*Genentech, Inc.*), Gerry Gitlin (*Biogen, Inc.*), Leo Gu (*Syntex Research*), Andy J. Jones (*Genentech, Inc.*), Victoria Knepp (*Alza, Corp.*), Leah Lipsich (*Boehringer Ingelheim*), Mike Mulkerrin (*Genentech, Inc.*), Rajiv Nayar (*Miles Laboratories*), John O'Connor (*Genentech, Inc.*), James "J.Q." Oeswein (*Genentech, Inc.*), Rodney Pearlman (*Megabios, Inc.*), Laurie Peltier (*Amylin*), Steve Prestrelski (*Alza Corp.*), Shelly Prince (*Univ. of Oklahoma*), Lynda Sanders (*Syntex Research*), Richard Senderoff (*Zymogenetics*), Paula Shadle (*SKB*), Mike Spellman (*Genentech, Inc.*), Robert Strickley (*Univ. of Utah*), Patricia Smialkowski (*SmithKline Beecham*), Glen Teshima (*Genentech, Inc.*), Jim Wells (*Genentech, Inc.*), and Tonie Wright (*Medical College of Virginia*).

REFERENCES

Abdel-Meguid, S. S., Shieh, H. S., Smith, N. W., Dayrenger, H. E., Violand, B. N., and Bentls, L. A., 1987, Three-dimensional structure of a genetically engineered variant of porcine growth hormone, *Proc. Natl. Acad. Sci. USA* **84(18)**:6434–6437.

Artigues, A., Birkett, A., and Schirch, V., 1990, Evidence for the in vivo deamidation and isomerization of an asparaginyl residue in cytosolic serine hydroxymethyltransferase, *J. Biol. Chem.* **265**:4853–4858.

Aswad, D. W., 1984, Stoichiometric methylation of porcine adrenocorticotropin by protein carboxyl methyl transferase requires deamidation of asparagine-25, *J. Biol. Chem.* **259**:10714–10721.

Battersby, J. E., Hancock, W. S., Canova-Davis, E., Oeswein, J. Q., and O'Connor, B., 1994, Diketopiperazine formation and N-terminal degradation in recombinant human growth hormone, *Int. J. Peptide Res.* **44**:215–222.

Bischoff, R., Lepage, P., Jaquinod, M., Cauet, G., Asker-Klein, M., Cleese, D., Laporte, M., Bayol, A., Dorsselaer, A. V., and Roitsch, C., 1993, Sequence specific deamidation: Isolation and biochemical characterization of succinimide intermediates or recombinant hirudin, *Biochemistry* **32**:725–734.

Blomback, B., 1967, Derivatives of glutamine in peptides, in: *Enzyme Structure*, Vol. II (C. H. W. Hirsh, ed.), Academic Press, New York, pp. 398–411.

Bodansky, M., Ondetti, M. A., Levine, S. D., and Williams, N., 1967, Synthesis of secretin. II. The stepwise approach, *J. Am. Chem. Soc.* **89**:6753–6757.

Bornstein, P., and Balian, G., 1970, The specific nonenzymatic cleavage of bovine ribonuclease with hydroxylamine, *J. Biol. Chem.* **245**:4854–4856.

Brange, J., Langkjaer, L., Havelund, S., and Vølund, A., 1992, Chemical stability of insulin. I. Hydrolytic degradation during storage of pharmaceutical preparations, *Pharm. Res.* **9**:715–726.

Brophy, R. T., and Lambert, W. J., 1994, The adsorption of insulinotropin to polymeric sterilizing filters, *J. Pharm. Sci. Tech.* **48**:92–94.

Cacia, J., Keck, R., Presta, L. G., and Frenz, J., 1996, Isomerization of an aspartic acid residue in the complementarity region of a recombinant antibody (in preparation).

Campbell, R. M., Lee Y., Rivier, J., Heimer, E. P., Felix, A. M., and Mowles, T. F., 1991, GRF analogs and fragments: correlation between receptor binding, activity and structure, *Peptides* **12**:569–574.

Canfield, R. E., 1963, The amino acid sequence of egg white lysozyme, *J. Biol. Chem.* **238**:2698–2707.

Canova-Davis, E., Baldonado, I. P., and Teshima, G. M., 1990, Characterization of chemically synthesized human relaxin by high performance liquid chromatography, *J. Chromatogr.* **508**:81–96.

Canova-Davis, E., Kessler, T. J., Lee, P., Fei, D. T. W., Griffin, P., Stults, J. T., Wade, J. D., and Rinderknecht, E., 1991, Use of recombinant DNA derived human relaxin to probe the structure of the native protein, *Biochemistry* **30**:6006–6013.

Capasso, S., Mazzarella, L., and Zagari, A., 1991, Deamidation via cyclic imide of asparaginyl peptides: dependence on salts, buffers and organic solvents, *Pept. Res.* **4**:234–238.

Chaudhary, V. K., Mizukami, T., Fuerst, T. R., Fitzgerald, D. J., Moss, B., Pastan, I., and Berger, E. A., 1988, Selective killing of HIV-infected cells by recombinant human CD4-*Pseudomonas* exotoxin hybrid protein, *Nature* **335**:369–372.

Chazin, W. J., Kördel, J., Thulin, E., Hofmann, T., Drakenberg, T., and Forsén, 1989, Identification of an isoaspartyl linkage formed upon deamidation of bovine calbindin D9k and structural characterization by 2D ^{1}H NMR, *Biochemistry* **28**:8646–8653.

Chou, F. C., Chou, C., Shaperia, R., and Kibler, R. F., 1976, Basis of microheterogeneity of myelin basic protein, *J. Biol. Chem.* **251**:2671–2679.

Cipolla, D., Gonda, I., Meserve, K. C., Weck, S., and Shire, S. J., 1994, Formulation and aerosol delivery of recombinant DNA-derived human deoxyribonuclease (rhDNase), in: *Formulation and Delivery of Proteins and Peptides* (J. L. Cleland and R. Langer, eds.), ACS Press, pp. 322–342.

Cipolla, D. C., and Shire, S. J., 1991, Analysis of oxidized human relaxin by reverse phase HPLC, mass spectrometry and bioassays, in: *Techniques in Protein Chemistry II*, (J. J. Villafranca, ed.), Academic Press, New York, pp. 543–555.

Cleland, J. L., Powell, M. F., and Shire, S. J., 1993, The development of stable protein formulations: a close look at protein aggregation, deamidation, and oxidation, *Crit. Rev. Ther. Drug Carrier Syst.* **10**:307–377.

Darrington, R. T., and Anderson, B. D., 1994, The role of intramolecular nucleophilic catalysis and the effects of self-association on the deamidation of human insulin at low pH, *Pharm. Res.* **11**:784–793.

Darrington, R. T., and Anderson, B. D., 1995, Evidence for a common intermediate in insulin deamidation and covalent dimer formation: Effects of pH and aniline trapping in dilute acidic solutions, *J. Pharm. Sci.* **84**:275–282.

Darrington, T., 1995, personal communication.

Daumy, G. O., Wilder, C. L., Merenda, J. M., McColl, A. S., Geoghegan, K. F., and Otterness, I. G., 1991, Reduction of biological activity of murine recombinant interleukin-1b by selective deamidation at asparagine 149, *FEBS* **278**:98–102.

DiAugustine, R. P., Gibson, B. W., Aberth, W., Kelly, M., Ferrua, C. M., Tomooka, Y., Brown, C. F., and Walker, M., 1987, Evidence for isoaspartyl[1] (deamidated) forms of mouse epidermal growth factor, *Anal. Biochem.* **165**:420–429.

Eisenberg, D., 1984, Three dimensional structure of membrane and surface proteins, *Annu. Rev. Biochem.* **53**:595–623.

Engelman, D. M., Steitz, T. A., and Goldman, A., 1986, Identifying nonpolar transbilayer helices in amino acid sequences of membrane proteins, *Annu. Rev. Biophys. Chem.* **15**:321–353.

Everett, R., 1995, personal communication.

Everett, R. R., Felice, C. J., Adomaitis, M., Plucinsky, M. C., and Siegel, R. C., 1995, Glutamate to pyroglutamate: Identification of a novel protein degradation mechanism, Ninth Symp. Protein Society.

Flatmark, T., 1966, On the heterogeneity of beef heart cytochrome c. III. A kinetic study of the non-enzymatic deamidation of the main subfractions (Cy I–Cy III), *Acta. Chim. Scand.* **20**:1487–1496.

Foster, L., personal communication, 1996.

Frenz, J., Shire, S. J., and Sliwkowski, M. B., Purified forms of DNase. US Patent 5,279,823 (1994).

Friedman, A. R., Ichhpurani, A. K., Brown, D. M., Hillman, R. M., Krabill, L. F., Martin, R. A., Zurcher-Neeley, H. A., and Guido, D. M., 1991, Degradation of growth hormone releasing factor analogs in neutral aqueous solution is related to deamidation of asparagine residues, *Int. J. Pept. Protein Res.* **37**:14–20.

Fukawa, H., 1967, Changes of glutamine-peptides on heating in aqueous media, *J. Chem. Soc. Jpn.* **88**:459–463.

Galletti, P., Iardino, P., Ingrosso, D., Manna, C., and Zappia, V., 1989, Enzymatic methyl esterification of a deamidated form of mouse epidermal growth factor, *Int. J. Pept. Protein Res.* **33**:397–402.

Geigert, J., Panschar, M., Forng, S., Huston, H. N., Wong, D. E., Wong, D. Y., Taforo, C., and Pemberton, M., 1988, The long term stability of recombinant (Serine-17) human interferon-b, *J. Interf. Res.* **8**:539–547.

Gitlin, G., Tsarbopoulos, A., Patel, S. T., Sydor, W., Pramanik, B. N., Jacobs, S., Westreich, L., Mittelman, S., and Bausch, J. N., 1996, Isolation and characterization of a monomethioninesulfoxide variant of interferon α-2b, *Pharm. Res.*, in press.

Grossenbacher, H., Marki, W., Coulot, M., Muller, D., and Richter, W. J., 1993, Characterization of succinimide-type dehydration products of recombinant products of recombinant hirudin variant 1 by electrospray tandem mass spectrometry, *Rapid Commun. Mass Spectrom.* **7**:1082–1085.

Gu, L. C., Erdos, E. A., Chiang, H., Calderwood, T., Tsai, K., Visor, G. C., Duffy, J., Hsu, W., and Foster, L. C., 1991, Stability of interleukin-1b (IL-1b) in aqueous solution: Analytical methods, kinetics, products and solution formulation implications, *Pharm. Res.* **8**:485–490.

Hageman, M. J., 1995, personal communication.

Hallahan, T. W., Shapiro, R., Strydom, D. J., and Vallee, B. L., 1992, Importance of asparagine-61 and asparagine-109 to the angiogenic activity of human angiogenin, *Biochemistry* **31**:8022–8029.

Halliwell, B., and Gutteridge, M. C., 1990, Role of free radicals and catalytic metal ions in human disease: An overview, *Methods Enzymol.* **186**:1–85.

Hamburger, R., Azaz, E., and Donbrow, M., 1975, Autooxidation of polyoxyethylenic non-ionic surfactants and of polyethylene glycols, *Pharm. Acta Helv.* **50**:10–17.

Harris, R. J., 1995, personal communication.

Harris, R. J., Kwong, M. Y., Molony, M. S., and Ling, V. L., 1995, *Mass Spectrometry in the Biological Sciences*, Humana Press.

Harris, R. J., Wagner, K. O., and Spellman, M. W. 1990, Structural characterization of a recombinant CD4-IgG hybrid molecule, *Eur. J. Biochem.* **194**:611–620.

Hayashi, T., Ohe, Y., and Hayahi, H., 1982, Human spleen histone H4. Isolation and amino acid sequence, *J. Biochem.* **92**:1995–2000.

Heller, D. L., and Qi, H., 1994, Stabilization of a 31 amino acid peptide formulation in a viscous dextran vehicle. AAPS 9th Natl. Meeting, San Diego, CA.

Henderson, L. E., Henriksson, D., and Nyman, P. O., 1976, Primary structure of human carbonic anhydrase C, *J. Biol. Chem.* **251**:5457–5463.

Hendy, G. N., Kronenberg, H. M., Potts Jr., J. T., and Rich, A., 1981, Nucleotide sequence of cloned cDNAs encoding human preproparathyroid hormone, *Proc. Natl. Acad. Sci. USA* **78**:7365–7369.

Herman, A., Sahakian, N., Taniguchi, G., Rohde, M., and Stoney, K., 1995, unpublished results.

Hershenson, S., Thompson, S., Luedke, E., Callahan, W., and Garcia, A., 1995, unpublished results.

Hopp, T. P., 1985, Prediction of protein surfaces and interaction sites from amino acid sequences, in: *Synthetic Peptides in Biology and Medicine* (K. Alitalo, P. Partanen, and A. Vaheri, eds.), Elsevier, Amsterdam, pp. 3–12.

Hopp, T. P., 1986, Protein surface analysis. Methods for identifying antigenic determinants and other interaction sites, *J. Immunol. Methods* **88**:1–18.

Hopp, T. P., and Woods, K. R., 1981, Prediction of protein antigenic determinants from amino acid sequences, *Proc. Natl. Acad. Sci. USA* **78**:3824–3828.

Hopp, T. P., and Woods, K. R., 1983, A computer program for predicting protein antigenic determinants, *Mol. Immunol.* **20**:483–489.

Hora, M., 1995, personal communication.

Hühmer, A. F. R., Gerber, N. C., Ortiz de Montellano, P. R., and Schöneich, C., 1996, Peroxynitrile reduction of calmodulin stimulation of neuronal nitric oxide synthase, *Chem. Res. Toxicol.* **9**: 484–491.

Inaba, M., Gupta, K. C., Kuwabara, M., Takahashi, T., Benz Jr., E. J., and Maede, Y., 1992, Deamidation of human erythrocyte protein 4.1: Possible role of aging, *Blood* **79**:3355–3361.

Ingram, R., and Warne, N., 1994, The stability of rhIL-11 as a function of pH, time, and temperature, AAPS 9th Natl. Meeting, San Diego, CA.

Johnson, B. A., Shirokawa, J. M., and Aswad, D. W., 1989a, Deamidation of calmodulin at neutral and alkaline pH: Quantitative relationships between ammonia loss and the susceptibility of calmodulin to modification by protein carboxyl methyltransferase, *Arch. Biochem. Biophys.* **268**:276–286.

Johnson, B. A., Shirokawa, J. M., Hancock, W. S., Spellman, M. W., Basa, L. J., and Aswad, D. W., 1989b, Formation of isoaspartate at two distinct site during in vitro aging of human growth hormone, *J. Biol. Chem.* **264**:14262–14271.

Johnson, D. M., and Gu, L. C., 1988, Absorption of drugs to bioavailability and bioequivalence, in: *Encyclopedia of Pharmaceutical Technology*, 1 (J. Swarbrick and J. C. Boylan, eds.), Marcel Dekker, New York, pp. 415–449.

Kanaya, S., and Uchida, T., 1986, Comparison of the primary sequences of ribonuclease U2 isoforms, *Biochem. J.* **240**:163–170.

Keck, R. C., 1995, Structural characterization of deamidated species formed during in vitro aging of recombinant interferon gamma, in preparation.

Kenigsberg, P., Fang, G., and Hager, L. P., 1987, Post-translational modifications of chloroperoxidase from *Caldariomyces fumago*, *Arch. Biochem. Biophys.* **254**:409–415.

Keutmann, H. T., Sauer, M. M., Hendy, G. N., O'Riordan, J. L. H., and Potts Jr., J. T., 1978, Complete amino acid sequence of human parathyroid hormone, *Biochemistry* **17**:5723–5729.

Keyt, B., and Cleland, J., 1995, personal communication.

Knepp, V. M., Whatley, J. L., Muchnik, A., and Calderwood, T. S., 1996, Identification of antioxidants for prevention of peroxide-mediated oxidation of recombinant human ciliary neurotrophic factor and recombinant human nerve growth factor, *J. Pharm. Sci. Tech.* **50**:163–171.

Kossiakoff, A. A., 1988, Tertiary structure is a principal determinant to protein deamidation, *Science* **240**:191–194.

Kroon, D. J., 1994, personal communication.

Kroon, D. J., Baldwin-Ferro, A., and Lalan, P., 1992, Identification of sites of degradation in a therapeutic monoclonal antibody by peptide mapping, *Pharm. Res.* **9**:1386–1393.

Kwong, M. Y., and Harris, R. J., 1985, Identification of succinimide sites in proteins by N-terminal sequence analysis after alkaline hydroxylamine cleavage, *Protein Sci.* **3**:147–149.

Kyte, J., and Doolittle, R., 1982, A simple method of displaying the hydropathic character of a protein, *J. Mol. Biol.* **157**:105–132.

Lauren, S. L., Arakawa, T., Stoney, K., and Rohde, M. F., 1993, Covalent dimerization of recombinant human interferon-γ, *Arch. Biochem. Biophys.* **306**:350–353.

Lim, A., Canova-Davis, E., Ling, V., Eng, M., Truong, L., Henzel, B., Stults, J., Harris, R., Gorrell, J., Heinsohn, H., McHugh, C., Weissburg, R. P., and Powell, M. F., 1996, Stability of rhTPO: Aggregation, oxidation, deamidation, diketopiperazine formation, AAPS Meeting, South San Francisco, CA, March, 1996.

Lin, L., Kunitani, M., and Hora, M., 1996, Interferon-b-1b (Betaseron): A model for hydrophobic therapeutic proteins, in: *Formulation, Characterization, and Stability of Protein Drugs: Case Histories* (R. Pearlman and Y. John Wang, eds.), Plenum Press, New York, pp. 275–301.

Maeda, H., and Kuromizu, K., 1977, Spontaneous deamidation of a protein antibiotic, neocarzinostatin, at weakly acidic pH, *J. Biochem.* **81**:25–35.

Maneri, L., 1994, personal communication.

McCrossin, L. E., Charman, W. N., Currie, G. J., and Charman, S. A., 1994, Degradation rates of Asn-99 and Cys-181–Cys-189 residues in native porcine growth hormone or the respective isolated tryptic peptides, AAPS 9th Natl. Meeting, San Diego, CA.

McGinity, J. W., Hill, J. A., and La Via, A. L., 1975, Influence of peroxide impurities in polyethylene glycols on drug stability, *Pharm. Sci.* **64**:356–357.

McKerrow, J. H., and Robinson, A. B., 1974, Primary sequence dependence on the deamidation of rabbit muscle aldolase, *Science* **183**:85.

Midelfort, C. F., and Mehler, A. H., 1972, Deamidation in vivo of an asparagine residue of rabbit muscle aldolase, *Proc. Natl. Acad. Sci. USA* **69**:1816–1819.

Nabuchi, Y., Fujiwara, E., Ueno, K., Kuboniwa, H., Asoh, Y., and Ushio, H., 1995, Oxidation of recombinant human parathyroid hormone: effect of oxidized position on the biological activity. *Pharm. Res.* **12**:2049.

Nguyen, T. H., Burnier, J., and Meng, W., 1993a, The kinetics of relaxin oxidation by hydrogen peroxide, *Pharm. Res.* **10**:1563–1571.

Nguyen, T. H., and Ward, C., 1993b, Stability characterization and formulation development of Alteplase, a recombinant tissue plasminogen activator, in: *Stability and Characterization of Protein and Peptide Drugs: Case Histories*, 5 (Y. J. Wang and R. Pearlman, eds.), Plenum Press, New York, pp. 91–134.

Nyberg, F., Bergman, P., Wide, L., and Roos, P., 1985, Stability studies on human pituitary prolactin, *Upsala J. Med. Sci.* **90**:265–277.

O'Kelley, P., Thomsen, L., Tilles, J. G., and Cesario, T., 1985, Inactivation of interferon by serum and synovial fluids, *Proc. Soc. Exp. Biol. Med.* **178**:407–411.

Oliyai, C., and Borchardt, R. T., 1993, Chemical pathways of peptide degradation. IV. Pathways, kinetics, and mechanism of degradation of an aspartyl residue in a model hexapeptide, *Pharm. Res.* **10**:95–102.

Ota, I. M., and Clarke, S., 1989, Enzymatic methylation of L-isoaspartyl residues derived from aspartyl residues in affinity-purified calmodulin. The role of conformational flexibility in spontaneous isoaspartyl formation, *J. Biol. Chem.* **264**:54–60.

Ota, I. M., Ding, L., and Clarke, S., 1987, Methylation at specific altered aspartyl and asparaginyl residues in glucagon by the erythrocyte protein carboxyl methyltransferase, *J. Biol. Chem.* **262**:8522–8531.

Paranandi, M. V., Guzzetta, A. W., Hancock, W. S., and Aswad, D. W., 1994, Deamidation and isoaspartate formation during in vitro aging of recombinant tissue plasminogen activator, *J. Biol. Chem.* **269**: 243–253.

Patel, K., 1993, Stability of adrenocorticotropic hormone (ACTH) and pathways of deamidation of asparaginyl residue in hexapeptide segments, in: *Stability and Characterization of Protein and Peptide Drugs: Case Histories*, (Y. J. Wang and R. Pearlman, eds.), Plenum Press, New York, pp. 201–220.

Patel, K., and Borchardt, R. T., 1990, Chemical pathways of peptide degradation. II. Kinetics of deamidation of an asparaginyl residue in a model hexapeptide, *Pharm. Res.* **7**:703–711.

Patel, S. T., and Gitlin, G., 1995, personal communication.

Pearlman, R., and Nguyen, T., 1992, Pharmaceutics of protein drugs, *J. Pharm. Pharmacol.* **44** (Suppl 1):178–185.

Poulter, L., Green, B. N., Kaur, S., and Burlingame, A. L., 1990, The characterization of native and recombinant proteins by electrospray mass spectrometry, in: *Biological Mass Spectrometry* (A. L. Burlingame and J. A. McCloskey, eds.), Elsevier, Amsterdam, pp. 119–128.

Presta, L. G., Lahr, S. J., Shields, R. L., Porter, J. P., Gorman, C. M., Fendley, B. M., and Jardieu, P. M., 1993, Humanizatiion of an antibody directed against IgE, *J. Immunol.* **151**:2623–2632.

Ragone, R., Facchiano, F., Facchiano, A., Facchiano, A. M., and Colonna, G., 1989, Flexibility plot of proteins, *Protein Eng.* **2**:497–504.

Robinson, A. B., and Rudd, C. J., 1974, *Deamidation of Glutaminyl and Asparaginyl Residues in Peptides and Proteins*, Academic Press, New York.

Robinson, A. B., Scotchler, J. W., and McKerrow, J. H., 1973a, Rates of nonenzymatic deamidation of glutaminyl and asparaginyl residues in pentapeptides, *J. Am. Chem. Soc.* **95**:8156–8159.

Robinson, A. B., and Tedro, S., 1973b, Sequence dependent deamidation rates for model peptides of hen egg, *Int. J. Pept. Protein Res.* **5**:275–278.

Sasaoki, K., Hiroshima, T., Kusumoto, S., and Nishi, K., 1989, Oxidation of methionine residues of recombinant human interleukin 2 in aqueous solutions, *Chem. Pharm. Bull. (Tokyo)* **37**:2160–2164.

Sasaoki, K., Hiroshima, T., Kusumoto, S., and Nishi, K., 1992, Deamidation at asparagine-88 in recombinant human interleukin-2, *Chem. Pharm. Bull. (Tokyo)* **40**:976–980.

Schultz, J., 1967, Cleavage at aspartic acid, *Methods Enzymol.* **11**:255–263.

Seid-Akhavan, M., Winter, W. P., Abramson, R. K., and Rucknagel, D. L., 1976, Hemoglobin Wayne: A frameshift mutation detected in human hemoglobin alpha chains, *Proc. Natl. Acad. Sci. USA* **73**: 882–886.

Senderoff, R. I., Wootton, S. C., Boctor, A. M., Chen, T. M., Giordani, A. B., Julian, T. N., and Radebaugh, G. W., 1994, Aqueous stability of human epidermal growth factor 1-48, *Pharm. Res.* **11**:1712–1720.

Sepetov, N. F., Krymsky, M. A., Ovchinnikov, M. V., Bespalava, Z. D., Isakova, O. L., Soucek, M., and Lebl, M., 1991, Rearrangement, racemization and decomposition of peptides in aqueous solution, *Pept. Res.* **4**:308–313.

Shahrokh, Z., Eberlein, G., Buckley, D., Paranandi, M. V., Aswad, D. W., Stratton, P., Mischak, R., and Wang, Y. J., 1994, Major degradation products of basic fibroblast growth factor: Detection of succinamide and iso-aspartate in place of aspartate-15, *Pharm. Res.* **11**:936–944.

Shen, F. J., Kwong, M., Keck, R. G., and Harris, R. J., 1996, Application of t-butylhydroperoxide to study sites of potential methionine oxidation in recombinant antibodies, in: *Techniques in Protein Chemistry*, Vol. 7 (D. Mershak, ed.), Academic Press, San Diego.

Shire, S. J., 1995, personal communication.

Sletten, K., Marhaug, G., and Husby, G., 1983, The covalent structure of amyloid-related serum protein SAA from two patients with inflammatory disease, *Z. Physiol. Chem.* **364**:1039–1046.

Son, K., and Kwon, C., 1995, Stabilization of human epidermal growth factor (hEGF) in aqueous solution, *Pharm. Res.* **12**:451–454.

Stadtman, E. R., 1990, Metal ion catalyzed oxidation of proteins: biochemical mechanism and biological consequences, *Free Radical Biol. Med.* **9**:315–325.

Steinberg, S. M., and Bada, J. L., 1983, Peptide decomposition in the neutral pH region via the formation of diketopiperazines, *J. Org. Chem.* **48**:2295–2298.

Stephenson, R. C., and Clarke, S., 1989, Succinimide formation from aspartyl and asparaginyl peptides as a model for the spontaneous degradation of proteins, *J. Biol. Chem.* **264**:6164– 6170.

Stevenson, C. L., Donlan, M. E., Freedman, A. R., Kubiak, T. M., and Borchardt, R. T., 1993, The effect of secondary structure on the rate of deamidation of several growth hormone releasing factor analogs, *Int. J. Pept. Protein Res.* **42**:497–503.

Sunahara, N., Kawata, S., Kaibe, K., Furuta, R., Yamayoshi, M., Yamada, M., and Kurooka, S., 1989, Differential determination of recombinant human interleukin-1α and its deamidated derivative by two sandwich enzyme immunoassays using monoclonal antibodies, *J. Immunol. Methods* **119**:75–82.

Svanti, J., and Milstein, C., 1972, The complete amino acid sequence of a mouse kappa light chain, *J. Biochem.* **128**:427–444.

Takao, T., Watanabe, H., and Shimonishi, Y., 1985, Facile identification of protein sequences by mass spectrometry, *Eur. J. Biochem.* **146**:503–508.

Teh, L., Murphy, L. J., Huq, N. L., Surus, A. S., Friesen, H. G., Lazarus, L., and Chapman, G. E., 1987, Methionine oxidation in human growth hormone and human chorionic somatomammotropin, *J. Biol. Chem.* **262**:6472–6477.

Teshima, G., and Canova-Davis, E., 1992, Separation of oxidized human growth hormone variants by reversed-phase high-performance chromatography, *J. Chromatogr.* **625**:207–215.

Teshima, G., Hancock, W. S., and Canova-Davis, E., 1995a, Effect of deamidation and isoaspartate formation on the activity of proteins, in: *Deamidation and Isoaspartate Formation in Peptides and Proteins* (D. Aswad, ed.), CRC Press, Boca Raton, FL, pp. 167–191.

Teshima, G., Porter, J., Yim, K., Ling, V., and Guzzetta, A., 1991a, Deamidation of soluble CD4 at asparagine-52 results in reduced binding capacity for the HIV-1 envelope glycoprotein gp120, *Biochemistry* **30**:3916–3922.

Teshima, G., Stults, J. T., Ling, V., and Canova-Davis, E., 1991b, Isolation and characterization of a succinimide variant of methionyl human growth hormone, *J. Biol. Chem.* **266**:13544–13547.

Teshima, G., and Wu, S., 1996, Capillary electrophoresis analysis of proteins, in: *Methods in Enzymology* (B. C. Kange and W. S. Hancock, eds.) in press.

Teshima, G., and Yim, K., 1995b, personal communication.

Tomizawa, H., Yamada, H., Ueda, T., and Imoto, T., 1994, Isolation and characterization of 101-succinimide lysozyme that possesses the cyclic imide at Asp-101-Gly102, *Biochemistry* **33**:8770–8778.

Tsuda, T., Uchiyama, M., Satao, T., Yoshino, H., Tsuchiya, Y., Ishikawa, S., Ohmae, S., Watanabe, S., and Miyake, Y. J., 1990, Mechanism and kinetics of secretion degradation in aqueous solutions, *J. Pharm. Sci.* **79**:223–227.

Tyler-Cross, R., and Schirch, V., 1991, Effects of amino acid sequence, buffers, and ionic strength on the rate and mechanism of deamidation of asparagine residues in small peptides, *J. Biol. Chem.* **266**: 22549–22556.

Venkatesh, Y. P., and Vithayathil, P. J., 1984, Isolation and characterization of monodeamidated derivatives of bovine pancreatic Ribonuclease A, *Int. J. Peptide Protein Res.* **23**:494–505.

Violand, B. N., Schlittler, M. R., Toren, P. C., and Siegel, N. R., 1990, Formation of isoaspartate 99 in bovine and porcine somatotropins, *J. Protein Chem.* **9**:109–117.

Violand, B. N., Siegel, M. R., Kolodziej, E. W., Toren, P. C., Cabonce, M. A., Siegel, N. R., Duffin, K. L., Zobel, J. F., Smith, C. E., and Tou, J. S., 1992, Isolation and characterization of porcine somatotropin containing a succinimide residue in place of aspartate 129, *Protein Sci.* **1**:1634–1641.

Volkin, D. B., and Middaugh, C. R., 1996, The characterization, stabilization, and formulation of acidic fibroblast growth factor, in *Formulation, Characterization, and Stability of Protein Drugs: Case Histories* (R. Pearlman and Y. John Wang, eds.), Plenum Press, New York, pp. 181–217.

Voorter, C. E. M., Roersma, E. S., Bloemendal, H., and de Jong, W. W., 1987, Age-dependent deamidation of chicken αA-crystallin, *FEBS Lett.* **221**:249–252.

Wang, J., 1995, personal communication.

Wang, J. H., Yan, Y. W., Garett, T. P., Liu, J. H., Rodgers, D. W., Garlick, R. L., Tarr, G. E., Husain, Y., Reinherz, E. L., and Harrison, S. C., 1990, Atomic structure of a fragment of human CD4 containing two immunoglobulin-like domains, *Nature* **348**:411–418.

Watanabe, C., 1991, Genentech Sequence Analysis Programs, Genentech, Inc., San Francisco.

Wearne, S. J., and Creighton, T. E., 1989, Effect of protein conformation on rate of deamidation: Ribonuclease and its ligands, *Proteins Struct. Funct. Genet.* **5**:8–12.

Wilson, J. M., Landa, L. E., Kobayashi, R., and Kelley, W. N., 1982, Human hypoxanthine-guanine ribosyltransferase, *J. Biol. Chem.* **257**:14830–14834.

Wingfield, P. T., Mattaliano, R. J., McDonald, H. R., Craig, S., Clore, G. M., Gronenborn, A. M., and Schmeissner, U., 1987, Recombinant-derived interleukin-1α stabilized against specific deamidation, *Protein Eng.* **1**:413–417.

Wood, D. C., Salsgiver, W. J., Kasser, T. R., Lange, G. W., Rowold, E., Violand, B. N., Johnson, A., Leimgruber, R. M., Parr, G. R., Siegel, N. R., Kimack, N. M., Smith, C. E., Zobel, J. F., Ganguli, S. M., Garbow, J. R., Bild, G., and Krivi, G. G., 1989, Purification and characterization of pituitary bovine somatotropin, *J. Biol. Chem.* **264**:14741–14747.

Wright, C. S., and Raikhel, N., 1989, Sequence variability in three wheat germ agglutinin isolectins: products of multiple genes in polyploid wheat, *J. Mol. Evol.* **28**:327–336.

Wright, H. T., 1991a, Sequence and structure determinants of the nonenzymatic deamidation of asparagine and glutamine residues in proteins, *Protein Eng.* **4**:283–291.

Wright, H. T., 1991b, Nonenzymatic deamidation of asparaginyl and glutaminyl residues in proteins, *CRC Crit. Rev. Biochem. Mol. Biol.* **26**:1–52.

Yao, Y., Yin, D., Jas, G. S., and Kuczera, K., Williams, T. D., Schöneich, C., and Squier, T. C., 1996, Oxidative modification of carboxyl-terminal vicinal methionine in calmodulin by hydrogen peroxide inhibits calmodulin-dependent activation of the plasma membrane Ca-ATPase, *Biochemistry* **35**:2767–2787.

CHAPTER 5

Comparative Structures of Mammalian Interferons

K. C. ZOON and R. WETZEL

A. Introduction

Our knowledge of the structure of mammalian interferons has been limited in t
past, predominantly because only minute quantities were available for structure
studies. Advances in amino acid analysis and sequence determination of picomol
quantities of protein have permitted the acquisition of composition and partial s
quence data for several native human and mouse interferons. However, the majc
ity of information on the structure of interferon has been the direct result of recor
binant DNA (rDNA) technology. Not only has this application of genetic e
gineering provided amino acid sequence data for a number of human interferor
but has also allowed the isolation of sufficient quantities of human interferon f
other structural studies, e.g., disulfide bond analysis and circular dichroism spe
troscopy. Studies aimed at determining the composition and structure of the ca
bohydrate moiety of interferon are, of course, dependent upon the availability
naturally derived material, and thus have been more limited.

B. Purification and Characterization of Native Interferons

I. Human Interferons-α

1. Purification

A summary of the major procedures developed for the purification of native h
man interferons-α (HuIFN-α)[1] was reported recently (ZOON 1981). Some of t
most powerful steps include immunoabsorbant affinity chromatography using e
ther monoclonal or polyclonal antibodies, sodium dodecylsulfate polyacrylamic
gel electrophoresis (SDS PAGE), and high pressure liquid chromatograph
(HPLC). Multiple species of native HuIFN-α have been isolated from virus-ir
duced cultures of buffy coat (ZOON et al. 1982a; BERG and HERON 1980), Namalw
(ZOON et al. 1979; ALLEN and FANTES 1980), chronic myelogenous leukemia (RU
BINSTEIN et al. 1981), and KG-1 (D. HOBBS 1981, personal communication) cells.

---

1 We have attempted to use current recommended nomenclature (leukocyte=α, fibrc
blast=β, immune=γ) as much as possible in this review. The following exceptions wi
be found, however: cloned interferon genes or their products derived from the work c
GOEDDEL et al. (1981) are referred to as either IFN-αA, B, C, etc., or LeIF-A, B, C, etc
single subtype interferons originally purified at the protein level (RUBINSTEIN et al. 198
are designated as α, β, γ based on their high pressure liquid chromatography retentio
times; in this case, the use of Greek letters in both systems of nomenclature is accommc
dated by placing the designations of RUBINSTEIN et al. in parentheses. For instance, IFN
α(Le$_1$ β_1) is purified subtype β_1 of human interferon-α

Table 1. Amino acid compositions of several native human interferons

Amino acid	HuIFN-$\alpha(\alpha_1)$[a,b]	HuIFN-$\alpha(\beta_1)$[a,b]	HuIFN-$\alpha(\gamma_2)$[c,d]	HuIFN-α (Ly, 18,500 daltons)[d,e]	HuIFN-β[d,f]
Asx	14.9	12.5	14.4	14.9	18.9
Thr	8.3	9.7	10.4	8.0	6.8
Ser	9.9	11.2	8.4	10.7	10.5
Glx	21.9	22.6	27.2	27.3	27.0
Pro	6.6	5.7	5.2	9.0	2.7
Gly	5.5	5.4	5.4	10.7	7.8
Ala	9.1	8.0	8.8	11.0	10.0
Val	8.0	6.5	7.6	7.7	6.0
Met	3.9	5.3	4.2	4.0	2.9
Ile	8.0	7.0	8.7	6.9	9.0
Leu	19.4	19.9	21.8	17.8	20.4
Tyr	4.3	4.8	5.2	3.8	7.5
Phe	7.4	9.5	9.7	7.1	9.4
His	3.3	3.1	3.6	4.4	4.9
Lys	11.4	9.7	10.9	10.4	11:6
Arg	6.7	8.8	9.3	9.6	10.9
Cys	4.2	3.4	3.2	1.8	1.7

[a] Levy et al. (1981)
[b] Based on 155 amino acid residues total (including 2 tryptophan residues)
[c] Rubinstein et al. (1981)
[d] Based on 166 amino acid residues total
[e] Zoon et al. (1981)
[f] Knight et al. (1980)

2. Characterization

Purified native HuIFN-α have an apparent molecular weight range of 16,000–23,000 (Rubinstein et al. 1981). The amino acid compositions of these interferons show a great deal of similarity. Several examples are shown in Table 1. These HuIFN-α exhibit extensive amino acid sequence homology among themselves (Fig. 1) as well as to those derived from DNA technology (see Fig. 3).

It is noteworthy that three major species of HuIFN-α: HuIFN-$\alpha(\alpha_1)$, HuIFN-$\alpha(\alpha_2)$, and HuIFN-$\alpha(\beta_1)$, isolated from chronic myelogeneous leukemia cells appear to lack the ten COOH terminal amino acids predicted from cDNA sequences of a number of HuIFN-α (see Fig. 3) as well as the sequences of several native HuIFN-α (Fig. 1) (Levy et al. 1981). No alterations in the specific activity of these abbreviated interferons have been observed (Levy et al. 1981). In addition, the amino acid sequences of HuIFN-$\alpha(\alpha_2)$ and HuIFN-$\alpha(\beta_1)$ appear to be virtually identical to the sequence of one of the major rDNA-derived interferons, HuIFN-αA, or HuIFN-α_2. Interestingly, the multiple species of native HuIFN-α show a range of antiviral activity titers on a number of animal cell lines and, in addition, they exhibit different ratios of cell growth inhibition to antiviral activity (Evinger et al. 1981). Similar properties are observed for the rDNA-derived HuIFN-α (see Table 3).

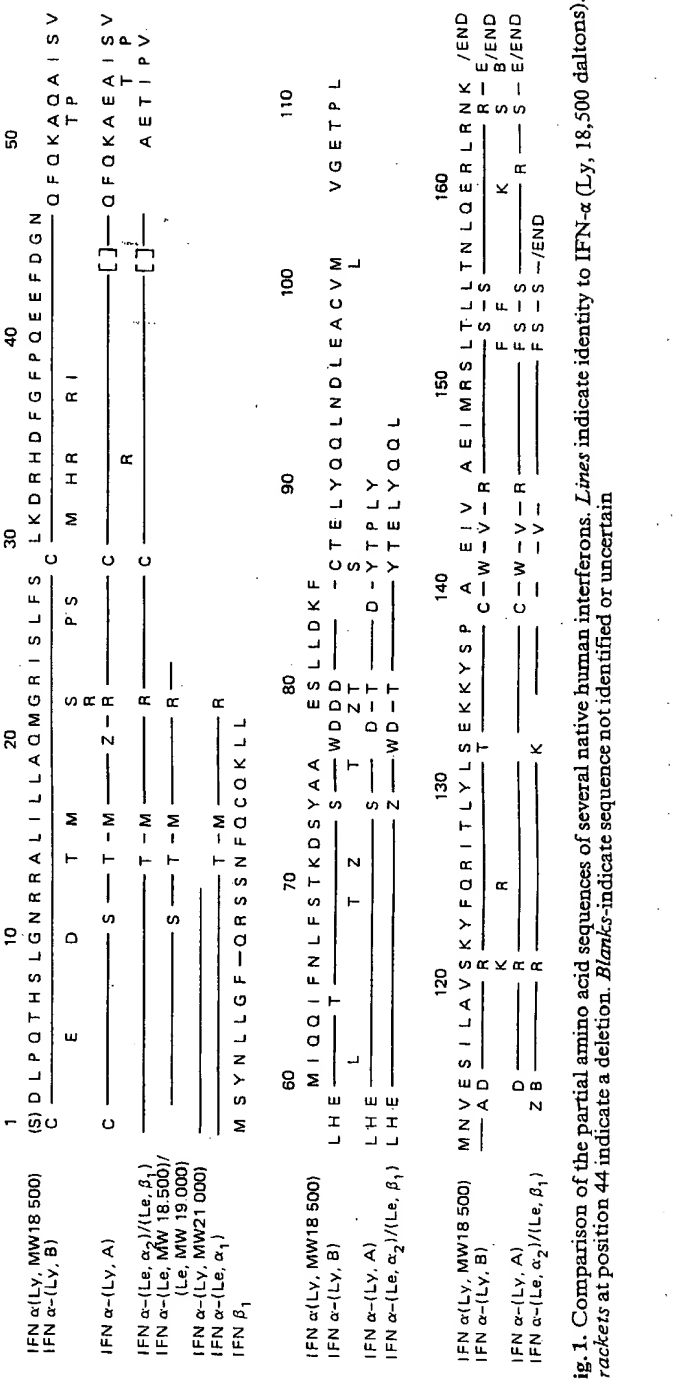


Fig. 1. Comparison of the partial amino acid sequences of several native human interferons. *Lines* indicate identity to IFN-α (Ly, 18,500 daltons). *Brackets* at position 44 indicate a deletion. *Blanks*-indicate sequence not identified or uncertain

Table 2. Amino acid composition of interferons from mouse Ehrlich ascites tumor cells (Cabrer et al. 1979)

	MuIFN-β (35,000 daltons) (residues/100 amino acids)	MuIFN-α (20,000 daltons) (residues/18,000 daltons)
Asx	9.3	14.2
Thr	7.7	9.4
Ser	4.8	9.2
Glx	16.4	23.9
Pro	2.8	5.7
Gly	3.9	6.3
Ala	5.3	9.8
Val	4.7	7.0
Met	3.5	4.2
Ile	4.4	5.2
Leu	11.7	16.7
Tyr	4.6	4.8
Phe	4.6	5.9
His	1.3	4.0
Lys	7.8	16.0
Arg	7.2	9.6
GLCN[a]	8.0	8.1
Cys	N.D.[b]	4.0

[a] GLCN = glucosamine
[b] N.D. = not determined

II. Human Interferon-β

1. Purification

Several purification schemes have been successfully developed to isolate human interferon-β (HuIFN-β) (Stewart 1981). Particularly noteworthy is the one-step purification procedure employing Blue Sepharose chromatography (Knight and Fahey 1981). At present only one species of biologically active HuIFN-β has been identified.

2. Characterization

Native HuIFN-β has an apparent molecular weight of 20,000. The amino acid composition is similar to that observed for the family of HuIFN-α (Table 1) and mouse interferons (Table 2). The partial amino acid sequence data obtained from microsequencing studies of the native protein (or proteins) is shown in Fig. 1 (Knight et al. 1980; E. Knight 1981, personal communication). This sequence is identical to that predicted from the nucleotide sequence of the HuIFN-β cDNA, excluding the signal peptide (see Fig. 4). Of the first 21 NH_2 terminal amino acids, only 1 residue at position 9 corresponds to that residue in the NH_2 terminal sequences of the native HuIFN-α.


```
                          1     5          10         15         20    2
MuIFN-β (MW 35 000)       I N Y K Q L Q L Q E R T N I R K ? Q E L L E Q L
        (MW 26 000)

MuIFN-α (MW 20 000)       A D L P Q T Y N L G N K G A L K V L A Q
```

Fig. 2. NH_2 terminal amino acid sequences of MuIFN-α and MuIFN-β

III. Human Interferon-γ

The purification of native human interferon-γ (HuIFN-γ) has been recently described (Yip et al. 1981). Two species with apparent molecular weights of 20,000 and 25,000 were detected by SDS PAGE (Yip et al. 1982). In contrast, with gel filtration, HuIFN-γ has an apparent molecular weight between 40,000 and 70,000. These studies suggest that native HuIFN-γ may be an aggregate. Currently, neither amino acid composition nor sequence data is available for native HuIFN-γ.

IV. Mouse Interferons

Three mouse interferons, one MuIFN-α and two MuIFN-β, have been purified to homogeneity (DeMaeyer-Guignard et al. 1978; Iwakura et al. 1978; Kawakita et al. 1978). In contrast to MuIFN-α and MuIFN-β, no homogeneous species of MuIFN-γ has been obtained. The amino acid compositions of MuIFN-β (molecular weight 35,000), MuIFN-β (molecular weight 26,000), and MuIFN-α (molecular weight 20,000) are shown in Table 2. Again a similarity is apparent among the MuIFN and between the MuIFN and the HuIFN-α and -β. Partial amino acid sequence data for MuIFN-α and both types of MuIFN-β are shown in Fig. 2 (Taira et al. 1980).

V. Comparison of Amino Acid Sequences of Human and Mouse Interferons

A comparison of the NH_2 terminal amino acid sequences of MuIFN-α and MuIFN-β to those of HuIFN-α and HuIFN-β clearly shows homology. Of the first 20 NH_2 terminal amino acids of native and rDNA-derived HuIFN-α, 8–13 are identical to those of MuIFN-α. Of the first 24 NH_2 terminal amino acid residues of HuIFN-β, 8 are identical to those of both forms of MuIFN-β.

C. Purification and Characterization of rDNA-Derived Interferons

I. Human Interferons-α

The DNA sequences of 13 distinct HuIFN-α cDNA clones indicate that these multiple HuIFN-α genes represent a family of homologous proteins. The primary

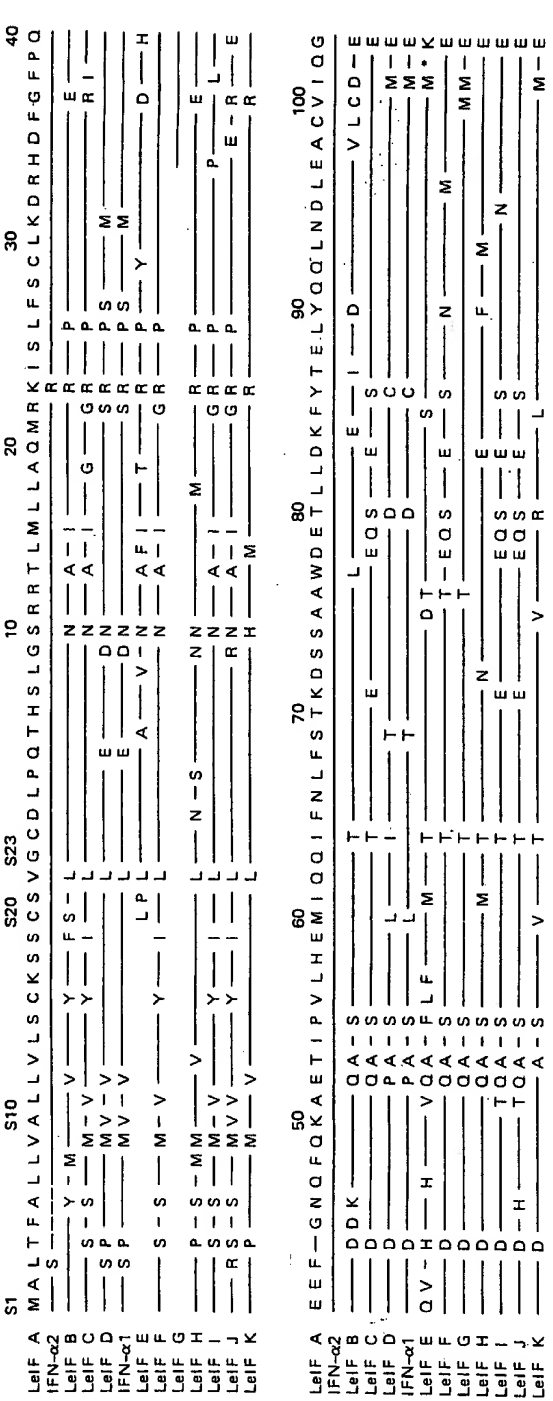


Fig. 3. Comparison of the protein sequences of 13 HuIFN-α deduced from nucleotide sequences. *Lines* indicate identity to HuIFN-αA (LeIF A). *Dash* for LeIF A indicates a deletion. *Asterisk* indicates inphase termination codons. S1 etc. denote signal peptide amino acid residues

amino acid sequences deduced from the DNA sequences of the clones are sho in Fig. 3 (Goeddel et al. 1981; Streuli et al. 1980). Each species consists of a amino acid residue signal peptide and a 165 (or 166) amino acid residue mat HuIFN-α protein, except for subtype E which appears to be a pseudogene c (Goeddel et al. 1981). Greater than or equal to 73% homology is observed for mature HuIFN-α amino acid sequences. Disregarding subtype E, approxima 60% of the amino acids are identical in all sequences (Goeddel et al. 1981). M of the amino acid changes can be attributed to single nucleotide alterations. predicted cysteine residues at positions 1, 29, 99 (or 100), and 139 of the mat HuIFN-α are highly conserved. HuIFN-αA and HuIFN-α_2 differ by only amino acid (Fig. 3). A single amino acid substitution was also observed HuIFN-αD and HuIFN-α_1 (Fig. 3). The signal peptide amino acid sequences sh >65% homology with 11 out of 23 positions identical. These signal peptic which are not found in isolated interferons, are presumably involved in cellular cretion, during which they are proteolytically removed. Some interferon genes h been expressed in yeast as well (Hitzeman et al. 1982). Genes coding for preint feron, when expressed in yeast, secrete correctly processed interferon into growth medium (Hitzeman et al., to be published).

Many of the native interferon structural genes have been engineered for pression in bacteria (for review, see Wetzel and Goeddel, to be published). *Es erichia coli*-derived HuIFN-αA has been purified to homogeneity using mo clonal antibody chromatography as the major purification step (Staehelin et 1981). The specific activity of the purified molecule is $1–3 \times 10^8$ U/per milligr protein (Staehelin et al. 1981; Wetzel et al. 1981) which is in the same range specific activities observed for native human interferons. The apparent molecu weight of purified subtype A is approximately 19,500 (Staehelin et al. 1981; W zel et al. 1981), again within the range observed for native human interferons. amino acid composition (Staehelin et al. 1981; Wetzel et al. 1981) of this spec is similar to those published for native human and mouse interferons-α and (Tables 1 and 2). The amino acid sequence of *E. coli*-derived HuIFN-αA is ide cal to that predicted by the nucleotide sequence of the gene for the mature prot (Fig. 3). This was determined by sequence analysis of the molecule's NH_2 termi (Wetzel et al. 1981) and of tryptic fragments collected from HPLC (Kohr a Wetzel, to be published). HPLC analysis of trypsin digests also allowed char terization of the disulfide bond arrangements in subtypes A (Wetzel 1981) anc (R. Wetzel 1982, unpublished work). Depending on fermentation conditions a on the strain of bacteria used, variable amounts of NH_2 terminal methionine observed. The rDNA-derived HuIFN-α like the native HuIFN-α exhibit a sp trum of antiviral activities on a variety of cell lines (Streuli et al. 1981; Weck al. 1981 a). In addition, genetically engineered hybrid HuIFN-α (see Sect. III have antiviral properties distinct from the parent molecules (Streuli et al. 19 Weck et al. 1981 b).

II. Human Interferons-β

In contrast to the multigene family of HuIFN-α, only a single HuIFN-β gene h been found (Derynck et al. 1980; Ohno and Taniguchi 1981). The amino acid

Fig. 4. Protein sequence of HuIFN-β deduced from nucleotide sequence. S1 etc. indicate signal peptide amino acid residues

quence deduced from the HuIFN-β gene sequence is shown in Fig. 4. This sequenc predicts, like the HuIFN-α cDNA sequences, a signal peptide (21 residues instea of the 23 residues observed for HuIFN-α) and 166 amino acid mature HuIFN- polypeptide. The signal peptide is absent from native HuIFN-β as was observe for native HuIFN-α. The 30 amino acids nearest the NH_2 terminus of mature IFN β expressed in *E. coli* have been determined and are as expected from the gene se quence (Harkins et al., to be published). The NH_2 terminal initiator methionin of HuIFN-β is about 80% removed by *E. coli*, giving a molecule one amino aci shorter than that isolated from fibroblast cell culture. No effect on interferon ac tivity of this difference has been observed.

III. Human Interferons-γ

Sequence analysis of the cloned cDNA coding for HuIFN-γ (Gray et al. 1982) al lows some insight into the structure of the protein molecule. This clone was iden tified as interferon-γ by the ability of derived DNA to command expression i mouse cells of an antiviral activity that could be neutralized by authentic anti-γ an tibodies, but not by anti-α or anti-β antibodies. In addition, other properties of the protein, predicted from the gene sequence, are consistent with those observed dur ing purification of lymphocyte-derived interferon-γ (see Sect. B.III). The amino acid sequence predicted for this molecule is shown in Fig. 5. The first 20 amino acids are probably a signal peptide for protein secretion, similar to sequences found in HuIFN-α and HuIFN-β genes. The mature protein coded on the gene is 146 amino acids long, approximately 20 amino acids shorter than HuIFN-α and HuIFN-β. Several groups have reported detection of limited homologies between HuIFN-γ and HuIFN-α or HuIFN-β (Gray et al. 1982; Gray and Goeddel 1982 Epstein 1982; DeGrado et al. 1982). The molecular weight, 17, 110, of the mole cule predicted by the DNA sequence is smaller than that reported for the activity derived from induced lymphocyte culture (Langford et al. 1979; DeLey et al 1980; Yip et al. 1981). This may be due to glycosylation and/or a dimeric nativ structure. There are two potential *N*-glycosylation sites in the predicted sequence at amino acid 28 and 100.

One striking feature of the amino acid sequence is the basicity of the molecule There are 27 basic amino acids (Arg + Lys) and only 19 acidic residues (Asp + Glu) Some of the excess positive charge may be neutralized in the glycosylated form by sialic acid, but the glycosylated, lymphocyte-derived form is basic as well, with a pI of about 8.7 (Yip et al. 1981). Eight of the basic residues occur in two cluster of four each (amino acids 89–92 and 131–134), similar to one of the cleavage site in the corticotropin–β-lipotropin precursor (Nakanishi et al. 1979).

Assuming the signal peptide is processed as expected (Gray et al. 1982), the mature HuIFN-γ molecule has only two cysteine residues, at positions 1 and 3 There are no proteins known which contain a disulfide bond between cysteines sep arated by only one amino acid (Richardson 1981). In addition, the activity of naturally derived HuIFN-γ is not sensitive to reducing agents tested (Yip et al. 1981). While the actual thiol structure of the unusual NH_2 terminal end of HuIFN- γ remains uncharacterized, there are clearly no disulfides in HuIFN-γ analogou

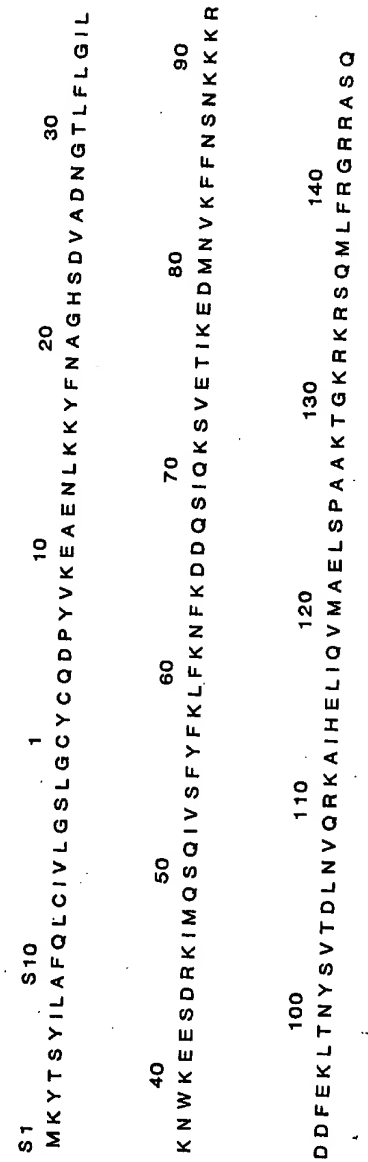


Fig. 5. Protein sequence of HuIFN-γ deduced from nucleotide sequence. S 1 etc. denote signal peptide amino acid residues. Gray et al. (1982)

to those found in HuIFN-α and presumed in HuIFN-β (Sect. D.I). Like interferons-α and -β, HuIFN-γ contains a large number of aromatic residues: ten phenylalanine, five tyrosine, and one tryptophan. A high α-helix content is expected based on structural prediction calculations (Sect. E).

D. Protein Structure and Interferon Activity

Until recently, structure–function studies on interferon suffered from: (a) the relatively low amount of protein available; (b) the lack of purity of most preparations; and (c) the heterogeneity of the preparations with respect to molecules possessing interferon activity. Recent work under these restrictions has centered on following extracted interferon activity by SDS PAGE. Limited proteolysis experiments (Otto et al. 1980; Braude et al. 1979, 1981 a) show that the apparent molecular weight of gel-extracted activity can be reduced when interferon preparations are exposed to some proteinases.

The availability of relatively large amounts of single subtypes of HuIFN-α (Wetzel et al. 1981; Staeheelin et al. 1981) as well as HuIFN-β (Harkins et al., to be published) has recently made possible structure–function studies on single molecular species (Wetzel et al. 1982). The following section includes preliminary results from some of the studies, as well as results taken from structure–function studies pursued at the DNA level.

I. Disulfide Bonds

There are four or five cysteines in the HuIFN-α, three in HuIFN-β, and two in HuIFN-γ. None of the naturally derived interferons, however, has been characterized for disulfide arrangements. It is known that while HuIFN-α (Mogensen and Cantell 1974) and HuIFN-β (Shepard et al. 1981) antiviral activity is sensitive to reducing agents, HuIFN-γ is insensitive (Yip et al. 1981).

Among the HuIFN-α, all cloned genes so far isolated contain conserved cysteines at positions 1, 29, 99, and 139 (numbering based on 166 amino acid length), which suggests two conserved disulfide bonds. Two disulfide bonds, between Cys-1 and Cys-98, and between Cys-29 and Cys-138, were characterized in HuIFN-αA synthesized in *E. coli* (Wetzel 1981; Wetzel et al. 1981). A similar arrangement is likely in HuIFN-αD (R. Wetzel 1982, unpublished work).

A chemical derivative of IFN-αA containing only the Cys-29–Cys-138 bond was found to possess full in vitro antiviral activity (Wetzel et al. 1982; Morehead et al., to be published). The lack of importance of the Cys-1–Cys-99 bond is also implied in the results of Streuli et al. (1980), who obtained active interferon from *E. coli* transformed with a plasmid containing an incomplete HuIFN-α_2 gene inserted into the β-lactamase gene of pBR322. If, as seems likely, their gene product is not a hybrid β-lactamase–interferon molecule, then it must arise from reinitiation of protein synthesis at the first interferon AUG in the "polycistronic" mRNA. Because the cDNA is incomplete at the 5′ end of the HuIFN-α gene, the first AUG occurs at amino acid 16 of the HuIFN-α molecule, and the isolated protein product thus can begin no earlier than Met-16. Thus, at least in HuIFN-α_2, the Cys-1–Cys-98 (or 99) disulfide as well as amino acids 1–15 seem to be nonessential or antiviral activity.

Previous work on crude IFN-α (Merigan et al. 1965; Mogenson and Cantell 1974) or cloned *E. coli* material (Stewart et al. 1980) has revealed differing

ses of IFN-α to reducing agent. Antiviral activity is either reversibly or irreversibly destroyed, depending on conditions and the IFN preparation used. Reduction of IFN-αA under native conditions inactivates the molecule and produces, depending upon reducing agent, varying amounts of disulfide-linked oligomers (Wetzel, to be published). Such thermally denatured preparations can be reactivated by a denaturation/renaturation cycle (using guanidine hydrochloride, urea, or sodium dodecylsulfate) followed by thiol–disulfide interchange or air oxidation.

This behavior of IFN-αA has been further studied using an *S*-sulfonate derivative of IFN-αA (Wetzel et al. 1982; Morehead et al., to be published). While this inactive disulfide-free derivative retains immunochemical relatedness to IFN-αA as well as the ability to regain antiviral activity after thiol–disulfide interchange, both these properties are lost after incubation of the derivative at 37 °C under native conditions. This denaturation was shown to be driven by a conformational change to a monomeric from of lower free energy. Exposure of this form to denaturants, followed by dialysis, regenerates the "proactive" conformation. This suggests that the inability of IFN-αA to survive reduction is due to the fact the the initial conformation of reduced IFN-αA decays at 37 °C (the minimum temperature for complete reduction) to a form which is incapable, under native conditions, of recovering an active or proactive conformation. The 29–138 disulfide, which is required for antiviral activity, is thus also important in maintaining IFN-αA in a critical conformation. In its absence, IFN-αA is subject to irreversible thermal denaturation (Wetzel, to be published).

II. Physical Studies

1. rDNA-Derived Interferons

a) Human Interferons-α

Preliminary investigation of HuIFN-αA by circular dichroism and ultraviolet spectroscopy indicates that the molecule is a typical globular protein with a densely packed, hydrophobic core. One can measure an α-helix content at neutral pH ranging from 45% to 70% (Bewley et al. 1982; M. Boublik and H. Kung 1982, personal communication), while no major β-sheet structure is apparent. Raman spectroscopy of IFN-αA gave these approximate values: α-helix, 49%; disordered helix, 25%; extended β-sheet, 8%; turns, 10% (R. Wetzel and R. Williams 1982, unpublished results). At least one of the molecule's tryptophans is tightly held in an asymmetric environment. This interaction, as well as about 50% of the α-helix, is reversibly lost on titration of the molecule to pH 2 (Bewley et al. 1982).

Ultracentrifugation studies on IFN-αA show a concentration-dependent aggregation in the neutral range, with sedimentation coefficients consistent with a dimeric or trimeric structure. The molecule behaves as a monomer at pH 2 and at lower concentrations (Shire 1982). A major conformational change at low pH can also be detected when IFN-αA is studied by pH titration. The change occurs around pH 3 and is entirely reversible. Several residues (Lys or Tyr) were found to ionize at abnormally low pH (Shire 1982).

b) Human Interferon-β

Circular dichroism studies on HuIFN-β purified from *E. coli* show it to contai about 55% α-helix (M. Boublik and H. Kung 1982, personal communication consistent with structure predictions (Fig. 7).

2. Interferon Fragments

COOH terminal fragments of HuIFN-α_1 containing residues 121–166 and 111–16 have been chemically-synthesized using solid-phase synthesis techniques (Arn heiter et al. 1981; Smith et al. 1981). These fragments do not exhibit antiviral ac tivity nor do they compete with radiolabeled native HuIFN-α for its binding sit on bovine kidney cells (Zoon et al. 1982b). However, they are antigenically cross reactive with HuIFN-α_1 and exhibit secondary structure as observed by circula dichroism studies. Fragment 111–166 yielded values of 24% α-helix and 36% β sheet (Arnheiter et al. 1981). The fragment 121–166 exhibited an α-helix conten consistent with the predicted α-helix content of HuIFN-α average (Smith et al 1981). Trypsin and cyanogen bromide fragments of IFN-αA, individually or i mixtures, had no detectable antiviral or receptor binding activities (Wetzel et al 1982).

III. Effect of Sequence Changes on Activity

1. NH_2 Terminal Variations

The first 15 amino acids of IFN-α_2 are probably not essential for antiviral activity (see Sect. D.I).

2. COOH Terminal Variations

Some of the COOH terminal amino acids of the cDNA-predicted interferon-α amino acid sequence are not essential for antiviral activity. Levy et al. (1981) have characterized by microsequencing of tryptic fragments several active interferons-α isolated from cell culture and were unable to locate the ten amino acids nearest the COOH terminus of these molecules. An interferon isolated from limited proteolysis of HuIFN-αA, which lacks the 13 amino acids nearest the COOH terminus, has full in vitro antiviral activity (Wetzel et al. 1982). In addition , a short HuIFN-αA has been made by introducing an early stop codon in the cloned gene after position 154. The gene product synthesized in *E. coli* exhibits 30% of the specific activity of the full length subtype (Franke et al., to be published).

3. cDNA-Encoded Analogs

At least 14 subtypes of interferon-α have been cloned, leading to structure–function information derived from comparisons of specific activities with amino acid sequences in the different proteins (Streuli et al. 1980; Yelverton et al. 1981; Weck et al. 1981a). In addition, these cloned genes can in some bases be used to generate artifical subtypes, by making hybrid genes that encode new sequence variants. This can be done by in vitro recombination of gene fra

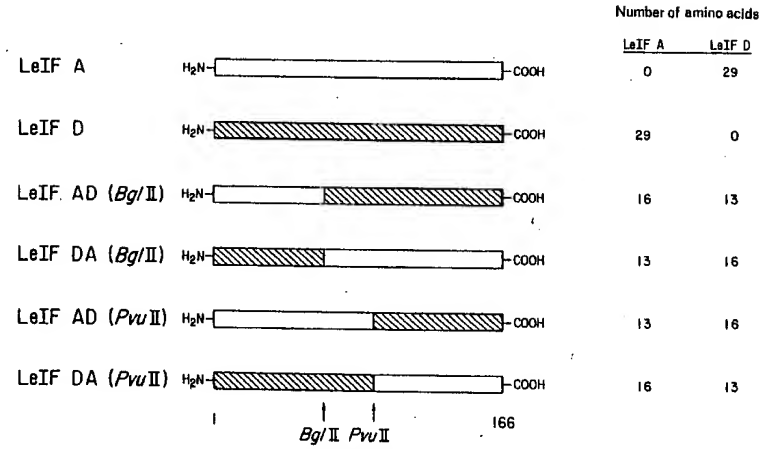


Fig. 6. Design of hybrid interferons-α produced by expression of cloned cDNAs derived from in vitro recombination of naturally derived cDNA. Weck et al. (1981 b)

Table 3. Relative specific activities of interferons

LeIF	Streuli et al. (1981)			Weck et al (1981 b)		
	MDBK	WISH	L929	MDBK	WISH	L929
A	100	100	3	100	100	6
D	131	14	100	74	15	100
AD (*Bgl*)	54	28	667	97	190	100,000
AD (*Pvu*)	46	144	333	47	280	4,000
DA (*Bgl*)	46	< 1	< 2	110	< 1	20
DA (*Pvu*)	46	< 1	< 1	110	< 2	100

a DNA restriction site common to two subtype genes. Using cloned cDNA for HuIFN-αA (α_2) and D (α_1), Streuli et al. (1981) and Weck et al. (1981 b) reported the construction of AD and DA hybrids and specific activity comparisons in a number of cell lines.

Figure 6 (Weck et al. 1981 b) shows the nature of these constructions at the *Bgl*II site (amino acid 61 of IFN-αA) and the *Pvu* site (amino acid 91 of IFN-αA). Despite the facts that the interferons produced in one study (Streuli et al. 1981) were synthesized at the ribosome as 21 amino acid NH_2 terminal extensions of the native interferons produced in the other study (Weck et al. 1981 b), and that the two groups used different techniques to compensate for possible differential stability in vivo of interferon analogs, the data produced are quite similar. Table 3 shows data from each group in which the analogs as well as the parent molecules were assayed in vitro on different cell lines.

Although some dramatic effects were observed on interferon activity on different cell lines, the data cannot be rationalized with respect to any simple structure–function model. Activity seems to be associated with one end of the molecule in one series and with the other end in another series (Streuli et al. 1981). Streuli et al. (1981) propose a model in which both ends of the molecule are involved in binding to receptor components capable of differentially responding to these NH. and COOH termini.

Hybrid experiments like these are valuable initial forays into interferon structure–function studies, and, aided by other physical or biochemical studies on the purified proteins, may yet provide real clues to the way the interferon molecule functions. In addition, some hybrid interferons may prove to be of clinical use. Nonetheless, it appears that higher resolution methods of analog generation, such as transpositions of shorter gene fragments, and ultimately site-specific mutagenesis, are more likely to elucidate interferon structure–function relationships. One such analog has already been made. By constructing a gene containing a Cys to Tyr mutation at position 141 of HuIFN-β, Shepard et al. (1981) demonstrated the importance of cysteine at this position. The lack of antiviral activity of this interferon analog may be due to its loss of ability to form a disulfide bond.

IV. Carbohydrate Content

1. Native Human Interferons-α

Studies designed to elucidate whether native HuIFN-α are in fact glycoproteins have yielded conflicting results (Grob and Chadha 1979; Bose et al. 1976; Allen and Fantes 1980). Recent amino acid composition and sequence studies of native and rDNA-derived HuIFN-α suggest HuIFN-α are not glycosylated. Amino acid analyses of native HuIFN-α do not show the presence of amino sugars (Allen and Fantes 1980). Amino acid sequence data of these interferons show the absence of an Asn-X-Ser(Thr) sequence, which is required for the glycosylation of asparagine residues. These observations do not omit the possibility of an *O*-glycosidic carbohydrate–peptide linkage. While it is controversial as to whether native HuIFN-α are glycosylated, it appears that the carbohydrate portion of the molecule is not necessary for biologic activity, since: (a) treatment of HuIFN-α with a glycosidase mixture from *Streptococcus pneumoniae* does not result in the loss of biologic activity (Bose et al. 1976); and (b) rDNA-derived *E. coli* HuIFN-α lacking carbohydrate have similar specific activities to native HuIFN-α.

2. Native Human Interferons-β

In contrast to HuIFN-α, HuIFN-β appears to be a glycoprotein. Amino acid analysis of native HuIFN-β indicates the presence of the amino sugars galactosamine and mannosamine (Tan et al. 1979), and amino acid sequence data obtained from rDNA-derived HuIFN-β shows a potential *N*-glycosidic linkage site at the asparagine in position 80 (see Fig. 4). It is important to note that the biologic specific activity of rDNA-derived HuIFN-β is similar to that of native HuIFN-β, thus again indicating the carbohydrate moiety is not essential for eliciting the activities of HuIFN-β. In addition, treatment of homogeneous native HuIFN-β with a glycosidase mixture results in an apparent molecular weight change of approximately 5,000 as observed by SDS PAGE (Knight and Fahey 1982). These studies suggest native HuIFN-β is a glycoprotein and that the carbohydrate portion is not essential for expression of its biologic activity.

3. Native Human Interferons-γ

Although HuIFN-γ has been purified to apparent homogeneity, no amino acid or amino sugar composition is available. The amino acid sequence deduced from the nucleotide sequence of the HuIFN-γ cDNA clone shows two potential *N*-glycosidic linkage points at asparagine residues 28 and 100 (see Fig. 5). In addition, chromatographic and inhibitor studies indicate HuIFN-γ are glycoproteins. They exhibit lectin specificity, i.e., they bind to concanavalin A (con A)–Sepharose and are eluted with α-methylmannopyranoside (Mizrahi 1978). Species of HuIFN-γ produced in the presence of tunicamycin, an inhibitor of the synthesis of *N*-acetylglucosaminylpyrophosphorylpolyisoprenol, do not bind to con A-Sepharose, but still exhibit antiviral activity (Mizrahi 1978).

4. Native Mouse Interferons

Many if not all MuIFNs (α, β, and γ) appear to be glycoproteins. Purified MuIFN-α and MuIFN-β stain with periodate Schiff's Reagent on SDS gels, indicating the presence of carbohydrate (DeMaeyer-Guignard et al. 1978). Two interferon species with apparent molecular weights of 15,000 and 18,000 are produced by NDV-induced mouse C243 cells in the presence of tunicamycin in lieu of the 24,000 daltons (MuIFN-α) and 35,000 daltons (MuIFN-β) species produced in the absence of the inhibitor (Raj and Pitha 1981). Experiments on con A-Sepharose binding suggest that MuIFN-α, MuIFN-β (Besancon and Bourgeade 1974), and MuIFN-γ (E. Havell 1982, personal communication) possess carbohydrate moieties. Changes in the isoelectric point of MuIFN-γ following neuramidinase treatment also support its glycoprotein nature (E. Havell 1982, personal communication). Amino acid analyses of MuIFN-α and MuIFN-β show the presence of the amino sugar, glucosamine (Table 2).

E. Structure Prediction

The predicted secondary structures for interferons shown in Figs. 7, 8 (Wetzel et al. 1982), and 9 (R. Wetzel 1982, unpublished work)-were calculated by the method of Garnier et al. (1978). Predictions by this method benefit in principle from the availability of a series of homologous protein sequences. Since it is based upon eight cDNA-predicted IFN-α amino acid sequences, the average IFN-α structure in Fig. 7 is thus expected to be significantly more reliable (5%–10%) than predictions of any individual interferon-α. Secondary structure calculations using the Chou-Fasman (1974) method for HuIFN-β_1 and HuIFN-αD have been published (Hayes 1980).

All the interferons are predicted to be highly helical (50%–70%) by either method. Available experimental data (Sect. D.II) for HuIFN-α and HuIFN-β is in agreement with these calculations. While overall helical contents were confirmed by experiment, the calculations are limited in their ability to locate elements of structure precisely. Other predictive or experimental methods must be used to refine the calculations.

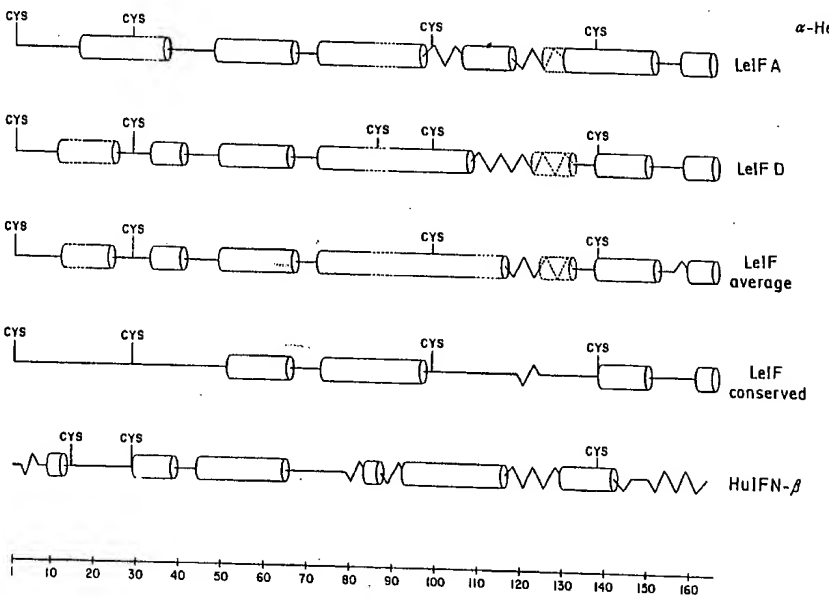


Fig. 7. Secondary structure predictions for HuIFN-α and HuIFN-β applying the algorithm of Garnier et al. (1978) to amino acid sequences predicted from sequences of cloned cDNA Residues were scored for their relative tendencies to exist in four possible states (α-helix, extended chain, reverse turn, and coil) based on values for each amino acid obtained by examination of 26 protein crystal structures. Only α-helix (*barrels*) and extended chain (β sheet, *zigzags*) are shown since they are predicted most accurately. Stretches equally likely to be in α-helix or extended chain are shown with these structures *dotted and superimposed* Regions with moderate helical potential which might be strengthened by adjoining helical regions are shown as *dotted connections* between helices

Figure 7 shows that HuIFN-αA and HuIFN-αD, differing in amino acid sequence, are predicted to have some structural homology, but also some differences The "average" structure shown is a best guess at HuIFN-α secondary structure which is assumed to be constant throughout the subtypes. The "conserved" HuIFN-α structure shows the strongest predicted structural elements, which are consistently predicted in all subtypes. One method of further refining the average prediction is to use a different algorithm to predict folding. Figure 8 shows a hydrophilicity profile for IFN-αA. These calculated affinities of polypeptide segments for an aqueous environment should be highest (most negative free energy) at solvent-exposed regions such as β-turns. The four reverse turns predicted by the algorithm of Garnier et al. (1978) are supported by their coincidence with maxima in the hydrophilicity curve. Predicted regions of flexible, solvent-exposed polypeptide can also be tested experimentally by limited proteolysis experiments. The arrows of Fig. 8 indicate points on the polypeptide chain cleaved by limited digestion with a variety of endoproteases. Cleavages at positions 7, 28, and 152 support the predictions, while cleavages at positions 103, 109, and 117 cast doubt on a strong α-helical character in this region of IFN-αA

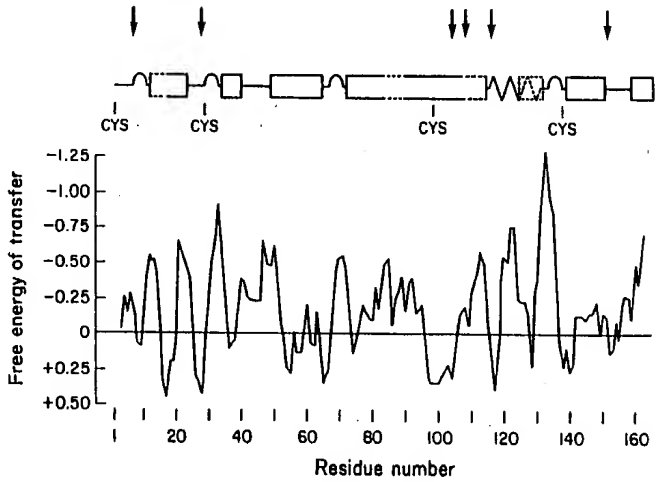


Fig. 8. Tests of the predicted average IFN-α structure. The average IFN-α structure from Fig. 7 is redrawn, including the four strongly predicted reverse turns, shown as loops. The hydrophilicity profile was computed using as a data base the free energies of transfer for the 20 amino acids determined by Janin (1979) from static accessibility tests on amino acids in 22 protein crystal structures. In jumps of one amino acid, average free energies along the primary sequence of IFN-αA were calculated for a moving window of five amino acids, and the values plotted with respect to the central amino acid. The Janin (1979) values were chosen based on their agreement with experimentally determined hydration potentials for amino acid side chains (Wolfenden 1981). The *arrows* on the top of the figure indicate loci along the primary sequence which suffered initial nicks by a variety of endoproteases under limiting conditions. W. Kohr and R. Wetzel (1981, unpublished work); Wetzel et al. (1982)

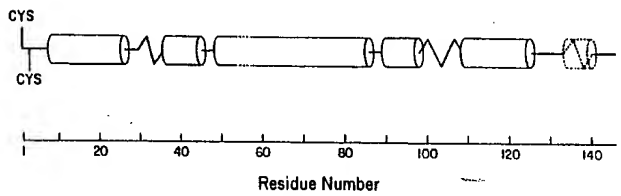


Fig. 9. Secondary structure prediction for HuIFN-γ by the algorithm of Garnier et al. (1978). See Fig. 7

Except for overall helix content, the HuIFN-β structure prediction very little resembles that for HuIFN-α (Fig. 7). This may be due in part to an error in the algorithm, to some real difference between interferon-α and -β structure, or may be a true prediction of a structural difference that is compensated in the real molecule by glycosylation. The potential *N*-glycosylation site of HuIFN-β at residue 80 is in fact in an area of the molecule which is predicted to have different structure from HuIFN-α. HuIFN-γ also predicted to contain over 50% α-helix (Fig. 9). While the prediction for IFN-α has had some experimental support (Sect. II.1.b), the predicted structures for IFN-β and IFN-γ have not been further tested.

Using the helical wheel analysis of Shiffer and Edmundson (1967), DeGr et al. (1982) have built a convincing case for significant *structural* homology tween interferons-α, -β, and -γ in a presumed α-helix in the middle portion of t primary sequences. They interpreted their results as evidence for divergent ev tion of interferons-α, -β, and -γ from a common precursor.

References

Allen G, Fantes KH (1980) A family of structural genes for human lymphoblastoid (le cyte-type) interferon. Nature 287:408–411

Arnheiter H, Thomas RM, Leist T, Fountoulakis M, Gutte B (1981) Physico-chemical antigenic properties of synthetic fragments of human leukocyte interferon. Na 294:278–280

Berg K, Heron I (1980) The complete purification of human leukocyte interferon. Sca Immunol 11:489–502

Besancon F, Bourgeade M (1974) Affinity of murine and human interferon for concanav A. J. Immunol 113:1061–1063

Bewley TA, Levine HL, Wetzel R (1982) Structural features of human leukocyte A as termined by circular dichroism spectroscopy. Int J Pept Protein Res 20:93–96

Bose S, Gurari-Rotman D, Ruegg UT, Corley L, Anfinsen CB (1976) Apparent dispens ity of the carbohydrate moiety of human interferon for antiviral activity. J Biol Cl 251:1659–1662

Braude IA, Lin LS, Stewart WE (1979) Differential inactivation of homologous and he ologous antiviral activity of human leukocyte interferon by a proteolytic enzyme. I chem Biophys Res Comm 89:612–619

Braude IA, Lin LS, Stewart WE (1981 a) Isolation of a biologically active fragment of man alpha interferon. J Interferon Res 1:245–252

Braude IA, Lin LS, Stewart WE (1981 b) Size characteristics of human leukocyte interfe under reducing and nonreducing conditions. Biochem J 193:947–951

Cabrer B, Taira H, Broeze RT, Kempe TD, Williams K, Slattery E, Konigsberg V Lengyel P (1979) Structural characteristics of interferons from mouse. Ehrlich asc tumor cells. J Biol Chem 254:3681–3684

Chou PY, Fasman GD (1974) Prediction of protein conformation. Biochemistry 13:222–

DeGrado WF, Wasserman ZR, Chowdhry V (1982) Sequence and structural homolo among type I and type II interferons. Nature 300:379–381

De Ley M, Van Damme J, Claeys H, Weening H, Heine JW, Billiau A, Vermylen C, Somer P (1980) Interferon induced in human leukocytes by mitogens: production, p tial purification and characterization. Eur J Immunol 10:877–883

DeMaeyer-Guignard J, Tovey MG, Gresser I, DeMaeyer E (1978) Purification of mo interferon by sequential affinity chromatography on poly(U)- and antibody-agar columns. Nature 271:622–625

Derynck R, Content J, DeClercq E, Volckaert G, Tavernier J, Devos R, Fiers W (19 Isolation and structure of a human fibroblast interferon gene. Nature 285:542–546

Epstein LB (1982) Interferon-gamma: success, structure, and speculation. Nature 295:4 454

Evinger M, Rubinstein M, Pestka S (1981) Antiproliferative and antiviral activities of l man leukocyte interferons. Arch Biochem Biophys 210:319–329

Franke AE, Shepard HM, Houck CM, Leung DW, Goeddel DV, Lawn RM (to be pu lished) Carboxy terminal region of hybrid interferons affects antiviral specificity

Garnier J, Osguthorpe DJ, Robson B (1978) Analysis of the accuracy and implications simple methods for prediciting the secondary structure of globular proteins. J Mol B 120:97–120

Goeddel DV, Leung DW, Dull TJ, Gross M, Lawn RM, McCandliss R, Seeburg PH, Ullrich A, Yelverton E, Gray PW (1981) The structure of eight distinct cloned human leukocyte interferon cDNAs. Nature 290:20–26

Gray PW, Goeddel DV (1982) Structure of the human immune interferon gene. Nature 298:859–863

Gray PW, Leung DW, Pennica D, Yelverton E, Najarian R, Simonsen CC, Derynck R, Sherwood PJ, Wallace DM, Berger SL, Levinson AD, Goeddel DV (1982) Expression of human immune interferon cDNA in E. coli and monkey cells. Nature 295:503–508

Grob PM, Chadha KC (1979) Separation of human leukocyte interferon components by concanavalin A-agarose affinity chromatography and their characterization. Biochemistry 18:5782–5786

Harkins RN, Weck PK, Apperson S, Haas P, Agarwal B (to be published) Structural and biological properties of purified bacteria derived human fibroblast interferon.

Hayes TG (1980) Chou-Faman analyses of the secondary structure of F and Le interferons. Biochem Biophys Res Commun 95:872–879

Hitzeman RA, Hagie FE, Levine HL, Goeddel DV, Ammerer G, Hall BD (1981) Expression of a human gene for interferon in yeast. Nature 293:717–722

Hitzeman RA, Leung DW, Perry LJ, Kohr WJ, Levine HL, Goeddel DV (to be published) Secretion of human interferons by yeast

Iwakura J, Yonehara S, Kawade Y (1978) Purification of mouse L cell interferon. J Biol Chem 253:5074–5079

Janin J (1979) Surface and inside volumes in globular proteins. Nature 277:491–492

Kawakita M, Cabrer B, Taira H, Rebello M, Slattery E, Weideli H, Lengyel P (1978) Purification of interferon from mouse Ehrlich ascites tumor cells. J Biol Chem 253:598–602

Knight E Jr, Fahey D (1981) Human fibroblast interferon: an improved purification. J Biol Chem 256:3609–3612

Knight E Jr, Fahey D (1982) Human interferon-β: effects of deglycosylation. J Interferon Res 2:421–429

Knight E Jr, Hunkapiller MW, Korant BD, Hardy RWF, Hood LE (1980) Human fibroblast interferon: amino acid analysis and amino terminal amino acid sequences. Science 207:525–526

Kohr W, Wetzel R (to be published) Amino acid sequence of E. coli-produced human alpha interferon subtype A by characterization of the HPLC/tryptic map.

Langford MP, Georgiades JA, Stanton GJ, Dianzani F, Johnson HM (1979) Large-scale production and physicochemical characterization of human immune interferon. Infect Immun 26:36–41

Levy WP, Rubinstein M, Shively J, Del Valle U, Lai CY, Moschera J, Brink L, Gerber L, Stern S, Pestka S (1981). Amino acid sequence of a human leukocyte interferon. Proc Natl Acad Sci USA 78:6186–6190

Merigan TC, Winget CA, Dixon CB (1965) Purification and characterization of vertebrate interferons. J Mol Biol 13:679–681

Mizrahi A, O'Malley JA, Carter WA, Takatsuki A, Tamura G, Sulkowski E (1978) Glycosylation of interferon: effects of tunicamycin on human immune interferon. J Biol Chem 253:7612–7615

Mogensen KE, Cantell K (1974) Human leukocyte interferon: a role for disulfide bonds. J Gen Virol 22:95–103

Morehead H, Johnston P, Wetzel R (to be published) The role of the disulfide bonds of human alpha interferon. I. Contribution to antiviral avtivity

Nakanishi S, Inoue A, Kita T, Nakamura M, Chang ACY, Cohen SN, Numa S (1979) Nucleotide sequences of cloned cDNA for bovine corticotropin-β-lipotropin prescursor. Nature 278:423–427

Ohno S, Taniguchi T (1981) Structure of a chromosomal gene for human interferon β. Proc Natl Acad Sci USA 78:5305–5309

Otto MJ, Sedmak JJ, Grossberg SE (1980) Enzymatic modifications of human fibroblast and leukocyte interferons. J Virol 35:390–399

Raj NBK, Pitha PM (1981) Interferon heterogeneity resulting from differences in glycosylation. J Interferon Res 1:595–600

Richardson JS (1981) The anatomy and taxonomy of protein structure. Adv Protein C 34:168–330

Rubinstein M, Levy WP, Moschera JA, Lai CY, Hershberg RD, Bartlett RT, Pest (1981) Human leukocyte interferon: isolation and characterization of several molec forms. Arch Biochem Biophys 210:307–318

Shepard HM, Leung D, Stebbing N, Goeddel DV (1981) A single amino acid change i man fibroblast interferon (IFN-β1) abolishes antiviral activity. Nature 294:563–565

Shiffer M, Edmundson AB (1967) Use of helical wheels to represent the structures of teins and to identify segments with helical potential. Biophys J 7:121–135

Shire S (1982) pH dependent behavior of purified biosynthetic human leukocyte interf A (LIF-A). Biophys J 37:384 a

Smith ME, Komoriya A, Anfinsen CB (1981) Chemical synthesis and properties of ments of human leukocyte interferon. In: DeMaeyer E, Galasso G, Schellekens H (The biology of the interferon system. Elsevier/North Holland, Amsterdam, pp 39–4

Staehelin T, Hobbs DS, Kung H, Lai CY, Pestka S (1981) Purification and characteriza of recombinant leukocyte interferon (IFLra) with monoclonal antibodies. J Biol C 256:9750–9754

Stewart WE II (ed) (1981) The interferon-system. Springer, Berlin Heidelberg New Y pp 168–171

Stewart WE II, Sarkar FH, Taira H, Hall A, Nagata S, Weissman C (1980) Compar of several biological and physicochemical properties of human leukocyte interfe produced by human leukocytes and by E. coli. Gene 11:181–186

Streuli M, Nagata S, Weissman C (1980) At least three human type α interferons: struc of α2. Science 209:1343–1347

Streuli M, Hall A, Boll W, Stewart WE II, Nagata S, Weissman C (1981) Target cell s ificity of two species of human interferon-α produced in Escherichia coli and of hy molecules derived from them. Proc Natl Acad Sci USA 78:2848–2852

Taira H, Broeze RJ, Jayaram GM, Lengyel P, Hunkapiller MW, Hood LE (1980) M interferons: amino terminal amino acid sequences of various species. Science 207: 530

Tan YH, Barakat F, Berthold W, Smith-Johannsen H, Tan C (1979) The isolation amino acid/sugar composition of human fibroblastoid interferon. J Biol C 254:8067–8073

Weck PK, Apperson S, May L, Stebbing N (1981 a) Comparison of the antiviral activ of various cloned human-α subtypes in mammalian cell cultures. J Gen Virol 57:233–

Weck PK, Apperson S, Stebbing N, Gray PW, Leung D, Shepard HM, Goeddel DV (1 b) Antiviral activities of hybrids of two major human leukocyte interferons. Nu Acids Res 9:6153–6166

Wetzel R (1981) Assignment of the disulfide bonds of leukocyte interferon. Nature 289: 607

Wetzel R (to be published) The role of the disulfide bonds of human alpha interferon Contribution to conformational stability

Wetzel R, Goeddel DV (1983) Synthesis of polypeptides by recombinant DNA meth In: Gross E, Meienhofer J (eds) The peptides: analysis, synthesis, biology. Acade New York (in press)

Wetzel R, Perry LJ, Estell DA, Lin N, Levine HL, Slinker B, Fields F, Ross MJ, Shi J (1981) Properties of a human alpha-interferon purified from E. coli extracts. J In feron Res 1:381–390

Wetzel R, Levine HL, Estell DA, Shire S, Finer-Moore J, Stroud R, Bewley TA (1 Structure-function studies on human alpha interferon. In: Merigan T, Friedman R, CF (eds) Chemistry and biology of interferons: relationship to therapeutics. U symposia 25. Academic, New York, pp 365–376

Wolfenden R, Andersson L, Cullis PM, Southgate CCB (1981) Affinities of amino acid chains for solvent water. Biochemistry 20:849–855

Yelverton E, Leung DW, Weck PK, Gray PW, Goeddel DV (1981) Bacterial synthesi a novel human leukocyte interferon. Nucleic Acids Res 9:721–741

Yip YK, Pang RHL, Urban C, Vilcek J (1981) Partial purification and characterization of human-γ (immune) interferon. Proc Natl Acad Sci USA 78:1601–1605
Yip YK, Barrowclough BS, Urban C, Vilcek J (1982) Molecular weight of human gamma interferon is similar to that of other human interferons. Science 215:411–413
Zoon K (1981) Purification, sequencing, and properties of human lymphoblastoid and leukocyte interferon. In: DeMaeyer E, Galasso G, Schellekens H (eds) The biology of the interferon system. Elsevier/North Holland, Amsterdam, pp 47–55
Zoon KC, Smith ME, Bridgen PJ, ZurNedden D, Anfinsen CB (1979) Purification and partial characterization of human lymphoblastoid interferon. Proc Natl Acad Sci USA 76:5601–5605
Zoon KC, Miller D, ZurNedden D, Hunkapiller M (1982a) Human leukocyte interferon: purification and amino terminal amino acid sequence of two components. J Interferon Res 2:253–260
Zoon KC, ZurNedden D, Arnheiter H (1982b) Specific binding of human alpha interferon to a high affinity cell surface binding site on bovine kidney cells. J Biol Chem 257:4695–4697

CHAPTER 6

Regulatory Control of Interferon Synthesis and Action

H. Smith-Johannsen, Y-T. Hou, X-Y. Liu, and Y-H. Tan

A. Regulatory Control of IFN Synthesis

I. Introduction

The induction of interferon (IFN) consists of a series of cellular events result in the transcription and translation of the IFN genes followed by events lead to the curtailment of these activities. After induction, cells remain refractory several hours before they can be stimulated for IFN synthesis again. Much co be learned of mammalian gene regulation from the mechanism or mechanisms which IFN genes are stimulated and regulated. This chapter reviews the evide regarding the number and nature of human IFN genes and the progress tow understanding the regulation of their expression.

II. Human IFN Genes

1. A Multigene Family

Several lines of evidence indicate that both IFN-α and IFN-β are families cons ing of a number of genes. Molecular cloning studies have revealed the existenc at least 15 gene-like sequences for IFN-α, including 5 pseudogenes (Nagata et 1980; Brack et al. 1981; Goeddel et al. 1981). None of the IFN-α genes so far amined contains introns (Nagata et al. 1980; Lawn et al. 1981a, b). Based on striction mapping, sequencing, and R-loop and heteroduplex analyses, Brack al. (1981) conclude that nine of the IFN-α genes are nonallelic and one is alle The data also indicate that some of the genes are clustered in linkage groups. C linkage group is 36 kilobases long and consists of three genes interspersed w three pseudogenes. A second group is 25 kilobases long and contains one gene a one pseudogene. The presence of extensive homologies in the flanking sequen of some of these genes, in particular the 35 kilobase linkage group, led Brack al. to speculate that these flanking regions may play a role in the regulation of expression of these IFN-α genes.

When human leukocytes are induced to produce IFN by Sendai virus, a hete geneous mixture of IFN-α mRNAs can be detected in the cytoplasmic extra (Sehgal et al. 1981). These mRNAs can be resolved into two size classes. The n jor population (IFN-α_S) corresponds to a size range of 0.8–1.4 kilobases an minor population (IFN-α_L) corresponds to a size range of 1.6–3.5 kilobases. T induction of the IFN-α_S mRNAs is preferentially inhibited by treatment of the duced leukocytes with 5,6-dichloro-1-D-ribofuranosylbenzimidazole (DRB). contrast, synthesis of IFN-α_L mRNAs appears to be increased in the presence

Interferons
and Their Applications

Contributors

J.A. Armstrong · A. Attallah · C. Baglioni · S. Baron · D.C. Burke
P.E. Came · W.A. Carter · E. De Clercq · V.G. Edy · E.C. Esber
K.H. Fantes · N.B. Finter · W.R. Fleischmann, Jr.
D.C. Gajdusek · J.A. Georgiades · C.J. Gibbs, Jr. · D. Gillespie
S.B. Greenberg · J.J. Greene · S.E. Grossberg · J.U. Gutterman
M.W. Harmon · N.O. Hill · H.E. Hopps · Y-T. Hou
P.M. Jameson · M. Krim · M. Langford · H.B. Levy
X-Y. Liu · P.I. Marcus · S. Panem · E. Pequignot · J.C. Petricciani
F.L. Riley · J.L. Sabran · A.D. Sagar · A.M. Salazar · D.S. Secher
J.J. Sedmak · P.B. Sehgal · R.A. Smith · H. Smith-Johannsen
G.J. Stanton · D.R. Strayer · D.A. Stringfellow · Y.H. Tan
J.L. Taylor · P.F. Torrence · P.O.P. Ts'o · P.K. Weck · D.A. Weigent
R. Wetzel · J.M. Zarling · K.C. Zoon

Editors

P. E. Came and W. A. Carter

30. Jan. 2004

Springer-Verlag
Berlin Heidelberg New York Tokyo 1984

CHAPTER 3

Evolution of Interferon Genes. D. GILLESPIE, E. PEQUIGNOT, and W. A. CARTER
With 11 Figures

CHAPTER 4

Comparative Analysis of Interferon Structural Genes
P. B. SEHGAL and A. D. SAGAR. With 2 Figures

CHAPTER 5

Comparative Structures of Mammalian Interferons
K. C. ZOON and R. WETZEL. With 9 Figures

CHAPTER 4

Comparative Analysis of Interferon Structural Genes

P. B. SEHGAL and A. D. SAGAR

A. Introduction

Interferons (IFNs) are a family of inducible proteins which exert potent biologic effects on target cells. These proteins render animal cells resistant to infection by a wide spectrum of viruses, inhibit cell proliferation and exert immunomodulatory effects on a variety of target cells (SEHGAL et al. 1982). IFNs can be induced in a large number of different animal species (STEWART 1979) and usually exert their effects on cells of homologous species. However, certain IFNs are also active on cells of heterologous species (STEWART 1979).

Human and murine IFNs are presently classified into α (leukocyte), β (fibroblast), and γ (immune) subtypes based primarily on their antigenic relationships. Thus, antisera raised against IFN-α, -β, or -γ do not cross-react with IFNs of a different type. However, antisera raised against human IFN-α do cross-react with certain species of murine IFN-α (STEWART and HAVELL 1980). This relationship between the IFN proteins also extends to the structure of the respective IFN genes. Thus human IFN-α cDNA sequences do not cross-hybridize human IFN-β or -γ-genes, but do cross-hybridize murine IFN-α genes (OWERBACH et al. 1981).

Recent advances in the characterization of IFN mRNA species, the molecular cloning of some of the corresponding cDNA molecules, and the elucidation of the structure of some of the human IFN genes have provided remarkable insights into the structural and evolutionary relationships that exist in this complex multigene family that codes for proteins which exert potent antiviral, anticellular, and immunomodulatory effects on animal cells. Some of the human IFN genes are closely related (cross-hybridize), others only distantly related (do not cross-hybridize); some are located in tandem on the same chromosome in the human genome, others are widely dispersed; some of the genes are coordinately expressed while others are expressed in a grossly noncoordinate manner. The structural and functional complexity of the human IFN gene family suggests that the induction of specific IFNs may represent finely tuned responses by different cells or tissues to particular physiologic or pathologic stimuli. The recent elucidation of the precise structural relationships between some of the human IFN genes represents a major advance in understanding the complex functions of this gene family.

B. Molecular Cloning of Some Human IFN-α cDNA and Chromosomal Genes

NAGATA and his colleagues (NAGATA et al. 1980a; STREULI et al. 1980) have described the molecular cloning of two distinct IFN-α cDNA species derived from

12 *S* polyadenylated RNA extracted from Sendai virus-induced human peripheral blood leukocytes. These two cDNA species, designated IFN-α_1 and IFN-α_2, were then used to screen a human DNA gene bank. This led to the isolation of at least ten distinct human IFN-α genes which cross-hybridize an α_1 cDNA probe (NAGATA et al. 1980b, 1981). Similarly, GOEDDEL and his colleagues isolated an IFN-α cDNA clone derived from 12 *S* polyadenylated RNA extracted from Sendai virus-induced human myeloblastoid cells (KG-1) (GOEDDEL et al. 1980a). This cDNA clone (LeIF A)[1] was then used as a DNA hybridization probe to isolate at least eight distinct, but cross-hybridizing cDNA clones from their 12 *S* mRNA library in pBR322 (GOEDDEL et al. 1980a, 1981). These investigators have also used these cDNA clones to screen a human DNA gene bank and have isolated up to 12 distinct, but cross-hybridizing IFN-α genes (LAWN et al. 1981a, b). This set of cross-hybridizing human IFN-α genes and their derived mRNAs and proteins has been designated IFN-α_S in order to distinguish it from a second set of human IFN-α_L mRNAs which do not appear to cross-hybridize IFN-α_S-specific DNA probes (SAGAR et al. 1981; SEHGAL et al. 1981a, b).

C. Molecular Cloning of a Human IFN-β cDNA and Its Chromosomal Gene

TANIGUCHI and his colleagues (TANIGUCHI et al. 1979, 1980a, b) were the first to report the molecular cloning of a single species of IFN-β cDNA derived from 12 *S* polyadenylated RNA extracted from poly(I)·poly(C)-induced diploid human fibroblasts. This species of cDNA is designated IFN-β_1 in order to distinguish it from other IFN-β mRNAs which do not appear to cross-hybridize an IFN-β_1 cDNA probe (SEHGAL and SAGAR 1980; WEISSENBACH et al. 1980; SAGAR et al. 1981, 1982). Numerous other investigators have also cloned and characterized IFN-β_1 cDNA (GOEDDEL et al. 1980b; DERYNCK et al. 1980a, b). Several investigators have screened human DNA gene banks using IFN-β_1 cDNA probes and have isolated and characterized a single gene corresponding to IFN-β_1 (HOUGHTON et al. 1981; TAVERNIER et al. 1981; DEGRAVE et al. 1981; LAWN et al. 1981c; OHNO and TANIGUCHI 1981; GROSS et al. 1981).

D. Comparative Structure of Some IFN-α and -β mRNAs and Proteins Deduced from cDNA Clones

The IFN-α cDNA clones described in Sect. B correspond to a group of cross-hybridizing mRNA species of length between 0.7 and 1.4 kilobases (SEHGAL et al. 1981a, b). This set of IFN-α mRNAs is collectively designated IFN-α_S. A second set of IFN-α_L mRNAs which corresponds to mRNA species of length between 1.6 and 3 kilobases (peak activity 1.8 kilobases) has not yet been cloned (SEHGAL et

1 IFN nomenclature is in a state of flux at the present time, with different laboratories using different designations. In the case of the human α system, LeIF A and LeIF D (GOEDDEL et al. 1981) are equivalent to HuIFN-α_2 and HuIFN-α_1 respectively (NAGATA et al. 1980a; STREULI et al. 1980)

al. 1981a, b). The IFN-β_1 cDNA described in Sect. C corresponds to an mRNA species of length approximately 0.9 kilobases (SEHGAL and SAGAR 1980). Although a total of five distinct human IFN-β mRNA species have been described recently (SEHGAL and SAGAR 1980; SAGAR et al. 1981, 1982) four of these have not yet been cloned. Thus, the discussion in Sect. D and E is restricted to the human IFN-α_S set and to the IFN-β_1 gene.

I. The Coding Regions

A detailed characterization of the 8–10 distinct IFN-α cDNA clones available at the present time has revealed that most of these would code for proteins which consist of 166 amino acids, except for IFN-α_2, which would code for a protein containing 165 amino acids (Fig. 1). There is approximately 80% homology in the amino acid sequences of the mature proteins, but 85%–95% homology in the DNA sequence in the coding region. Two domains, amino acids 28–80 and 115–150 are highly conserved in all of these IFN-α proteins. These regions may represent the biologically active sites on these proteins. Studies on the activity of hybrid interferons derived from fused cloned IFN cDNA (STREULI et al. 1981) where codons for the NH_2 terminal amino acids 63 or 92 of IFN-α_1 are fused with codons for the remainder of the COOH terminal amino acids of IFN-α_2 and vice versa suggest that species-specific IFN activity segregates with the NH_2 terminal portion of the IFN molecule. These data suggest that the region 28–80 may contain the site which binds to the cell surface receptor. It has been suggested that the second region (amino acids 115–150) may have a role in modulating this binding or may contain a site responsible for some other biologic function (STREULI et al. 1981).

IFN-β_1 cDNA also codes for a mature protein of 166 amino acids (TANIGUCHI et al. 1980a, c). IFN-β_1 is only 29% homologous with IFN-α_1 at the protein level but is ~45% homologous in the DNA sequence of the coding region (TANIGUCHI et al. 1980c). The two conserved domains in codons 28–80 and 115–150 observed in IFN-α_1 and -α_2 proteins are also conserved in IFN-β_1. However, the degree of nucleotide sequence conservation is not sufficient for cross-hybridization of IFN-β_1 RNA or DNA with IFN-α_1 DNA probes, even under relaxed hybridization conditions. Furthermore, the degree of amino acid sequence conservation is not sufficient for cross-reaction between antisera raised against IFN-α or -β and the heterologous interferons.

The IFN-protein sequences deduced from the cDNA clones indicate marked conservation of cysteine at positions 1, 29, 98 or 99, and 138 or 139 in the IFN-α proteins and the presence of cysteine residues at positions 31 and 141 in IFN-β (Fig. 1; STREULI et al. 1980; GOEDDEL et al. 1981; WETZEL 1981). In the IFN-α proteins, Cys-1 is bonded to Cys-98 or -99 and Cys-29 to Cys-138 or -139 by disulfide bridges (WETZEL 1981). Similarly, IFN-β_1 may contain Cys-31 bonded to Cys-141.

The natural IFN-α proteins characterized to date are devoid of carbohydrate moieties (RUBINSTEIN et al. 1981; ALLEN and FANTES 1980) whereas IFN-β_1 has been shown to be a glycoprotein (KNIGHT 1976; TAN et al. 1979). Attachment of carbohydrate through *N*-glycosidic linkage is known to occur on the asparagine in the triplets Asn-X-Ser or Asn-X-Thr and the presence of this sequence is a necessary, but not a sufficient condition for glycosylation (NEUBERGER 1972). IFN-β_1

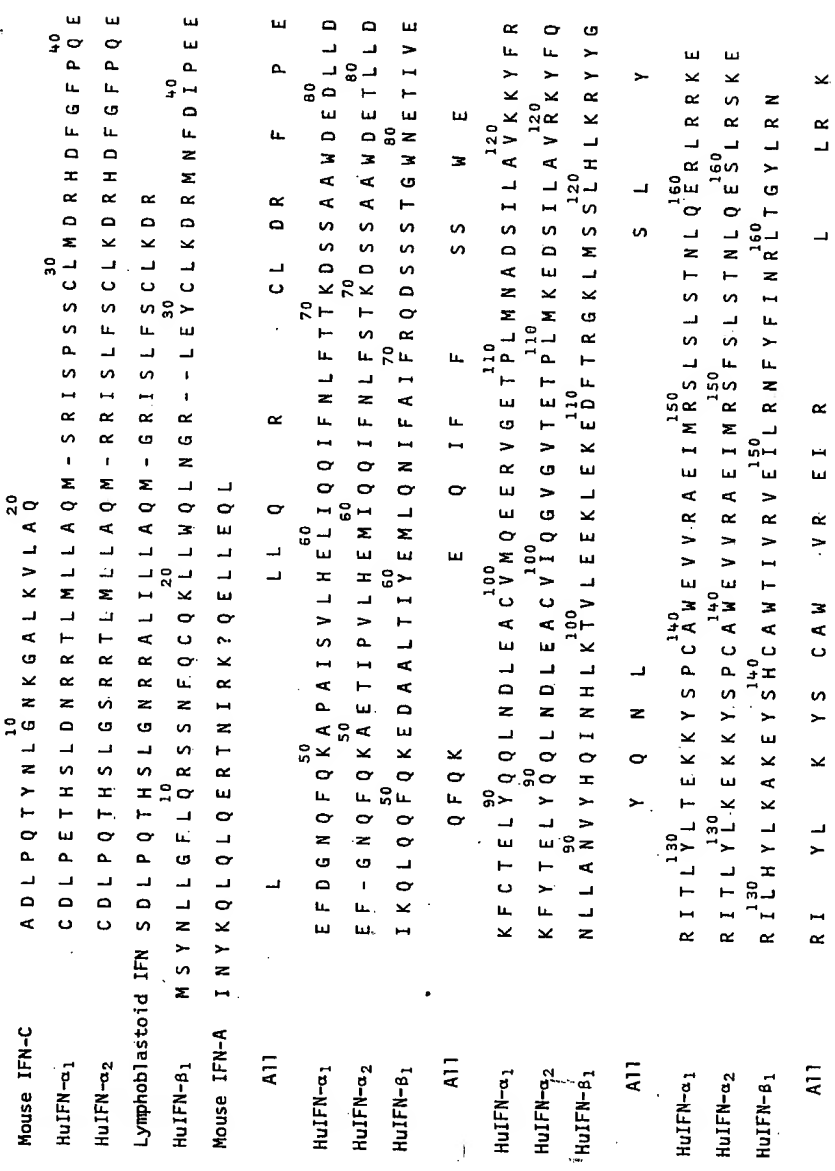


Fig. 1. The coding region of some IFN genes. Comparison of amino acid sequences of human IFN-α_1 and -α_2, deduced from the cDNA sequence (MANTEI et al. 1980), with human lymphoblastoid IFN, deduced from amino acid analyses, murine IFN-A and -C, deduced from NH_2 terminal amino acid analyses (TAIRA et al. 1980), and human fibroblast IFN-β_1 from the cDNA sequence (TANIGUCHI et al. 1980a). The amino acids are written according to the one-letter notation as recommended by the IUPAC-IUB Commission on Biochemical Nomenclature (DAYOFF 1978). A-alanine; C-cysteine; D-aspartic acid; E-glutamic acid; F-phenylalanine; G-glycine; H-histidine; I-isoleucine; K-lysine; M-methionine; N-asparagine; P-proline; Q-glutamine; R-arginine; S-serine, T-threonine; V-valine; W-tryptophan; Y-tyrosine. A *question mark* indicates a sequence not yet identified. *Dashes* have been introduced to obtain maximum homology. Mouse IFN-C corresponds to IFN-α and mouse IFN-A corresponds to IFN-β. STREULI et al. (1980)

contains such an asparagine at position 80 of the amino acid sequence while nc of the IFN-α protein sequences deduced to date indicates the presence of this ami acid triplet.

A number of cDNA molecules which have unusual features in their coding gions have been cloned:

1. The IFN-α cDNA designated "LeIFN E" (GOEDDEL et al. 1981) conta one nucleotide more than the other IFN-α cDNAs at position 187, correspond to amino acid 40. Subsequent to this translation termination condons are enco tered in all three phases. Thus "LeIFN E" cDNA appears to represent a pseu gene that is transcribed and expressed at a low level in virus-induced myeloblast cells.

2. Two IFN-α cDNA clones designated "LeIFN H" and "LeIFN H_1" are tually identical except for a single nucleotide deletion in H_1 (at nucleotide 545, which is followed six nucleotides later by a correcting insertion (after nucleot 551, G), thus restoring the original reading frame. These two clones probably r resent allelic genes (GOEDDEL et al. 1981).

3. An IFN-β_1 cDNA has been isolated in which the deduced protein seque includes the change of Cys-141 to Tyr-141. The protein expressed in *Escheric coli* corresponding to this unusual cDNA displays no antiviral activity, does compete with anti-IFN-β immunoglobulin, and does not bind to cell membra (SHEPARD et al. 1981). Thus, it appears that Cys-141 may play an important r in the biologic activity of IFN-β_1.

While many of the conclusions about these IFN-α and -β_1 proteins deduc from the nucleotide sequences of the cDNA clones have been confirmed by dir analyses of the mature proteins (KNIGHT et al. 1980; TANIGUCHI et al. 1980a. MAEDA et al. 1980; ALLEN and FANTES 1980), a recent report (LEVY et al. 1981) su gests that three particular species of mature IFN-α proteins derived from virus-duced human leukocyte cell cultures (the cells were obtained from patients w chronic myelogenous leukemia) lack the ten COOH terminal amino acids pred ed by the DNA sequence.

II. The Signal Peptides

IFNs are secretory proteins. Thus, each of the IFN cDNA clones that has b characterized reveals the presence of a hydrophobic signal peptide 21–23 ami acids long (TANIGUCHI et al. 1980a, c; MANTEI et al. 1980a; GOEDDEL et al. 198 b). There is a greater degree of variability in the signal sequence than in the cod regions. The six IFN-α signal peptides that have been deduced consist of 23 ami acids, are approximately 70% homologous to each other with 11 (43%) of amino acids completely conserved. On the other hand the 21 amino acid IFN signal peptide is markedly different from the IFN-α signal peptides (~60% div gence in nucleotide sequence) (TANIGUCHI et al. 1980c; STREULI et al. 1980).

III. The Noncoding Regions

The 5′ noncoding region in the IFN-α and β_1 mRNAs is approximately 65–75 cleotides long (HOUGHTON et al. 1981; NAGATA et al. 1980b; LAWN et al. 1981

	(e)	(d)	(c)	(b)	(a)	Corresponding designation
	-125	-80	-59	-30	-1	
1	TGAAAACCCATG	G-AAAGTA	TTTGGAA	TATTTAA	CA	IFN-α_2
2	TCTACACCCATG	GAAAAAAA	TTCAGAA	TATTTAA	CA	
3	CTTAAACACATG	G-AAAGTA	TTTAGAA	TATTTAA	CA	
4	TTTAAACACATG	G-AAAGTA	TTTAGAA	TATTTAA	CA	
5	TTTAAACACATG	G-AAAGTA	TTTAGAA	TATTTAA	CA	
6	TCTAAAATCATG	G-AAAGTG	TTAAGAA	TATTTAA	CA	IFN-α_1
7	TCTATACCCATG	G-AAAGTA	TTCTGAA	TATTTAA	CA	
	-114	-72	-57	-31	-1	
8	TACTAAAATG	GAAAGTGG	CTCTGAA	TATAAA	CA	IFN-β_1

Fig. 2a–e. The 5′ flanking region of some IFN genes. Comparison of homologous nucleotide sequences, upstream from the transcription initiation site of several human IFN-α genes and the IFN-β_1 gene. These sequences could conceivably play a role in gene regulation. a transcription initiation site; b Goldberg–Hogness box; c and e presumably involved in induction; d CCAAT box thought to be involved in RNA polymerase II binding. Sequences 1–7 taken from LAWN et al. (1981 b); 8 from DEGRAVE et al. (1981)

There is marked (~75%) sequence homology in the 5′ noncoding regions of seven of the IFN-α genes sequenced so far. On the other hand the 5′ noncoding region of IFN-β_1 has a much lower degree of homology with the IFN-α sequences. The 3′ noncoding regions of IFN-α and -β_1 mRNAs are highly variable. The length of the 3′ noncoding region can vary from 203 (IFN-β_1) to approximately 506 (LeIF A (+175)) nucleotides preceding the poly(A). There is approximately 50% homology in the nucleotide sequence in this region among the IFN-α mRNAs. Thus, DNA restriction fragments corresponding to this region can be used as hybridization probes for the individual IFN-α mRNA species (STREULI et al. 1980; GOEDDEL et al. 1981). There is only 30% homology between the nucleotide sequence in this region in IFN-β_1 mRNA and IFN-α_1 and -α_2 mRNAs (STREULI et al. 1980).

The hexanucleotide, AAUAAA precedes the site of polyadenylation in many eukaryotic cellular mRNAs by 15–25 nucleotides (PROUDFOOT 1976). IFN-β_1 mRNA contains the AAUAAA sequence 20 nucleotides internal to the poly(A) site. One-half of the IFN-α cDNAs sequenced contain the corresponding AATAAA sequence approximately 15–20 nucleotides from the poly(A) site (LeIFN A, D, F, and G; GOEDDEL et al. 1981) while several others (LeIFN B, C, E, and H; GOEDDEL et al. 1981) contain the related sequence ATTAAA. LeIFN B contains the ATTAAA approximately 400 nucleotides from the end of the coding region, but is not polyadenylated until after a second ATTAAA sequence is reached approximately 485 nucleotides into the 3′ noncoding region (GOEDDEL et al. 1981). Whereas LeIF A or IFN-α_2 is usually polyadenylated approximately 330 nucleotides from the end of the coding region and contains the AAUAAA sequence 20 nucleotides proximal to this poly(A) site (GOEDDEL et al. 1981), a variant

cDNA [LeIFN A (+175)] has been cloned which represents an mRNA species which polyadenylation did not occur at this site, but at a location 175 nucleotic further downstream (LAWN et al. 1981 b). A second AATAAA hexanucleotide p cedes the 3′ end of this extended cDNA clone (LAWN et al. 1981 b).

The length of the 3′ poly(A) tails present in IFN-α mRNA species has not be investigated. The length of the 3′ poly(A) in cytoplasmic IFN-β_1 mRNA has be estimated to be approximately 100 nucleotides and that of IFN-β_2 mRNA to approximately 200 nucleotides (SOREQ et al. 1981). Since the group of IFN-α_s ge and the IFN-β_1 gene are known to be devoid of introns (Sect. E) it is intriguing determine whether newly synthesized mRNAs derived from these genes have sho er poly(A) than newly synthesized mRNA molecules derived from most other c lular split genes (150–250 nucleotides). mRNAs derived from mammalian histo genes, which also lack introns, are devoid of 3′ poly(A). The degree of variabil in the 3′ noncoding sequence suggests that this region may not be particularly c cial to the translational function of IFN mRNAs. Indeed, deletion of the poly(and large (100–200 nucleotides) segments of the 3′ noncoding sequence internal the poly(A) does not affect the translational function of IFN-β_1 and -β_2 mRN in *Xenopus laevis* oocytes (SOREQ et al. 1981).

E. Comparative Structure of Some IFN-α and -β_1 Chromosomal Genes

A set of up to 12 distinct, but cross-hybridizing genes and pseudogenes has be isolated by screening human DNA gene banks (e.g., in lambda phage charon 4 using IFN-α cDNA probes (NAGATA et al. 1980 b, 1981; LAWN et al. 1981 a, b). contrast, a single IFN-β_1 gene has been isolated in this manner (HOUGHTON et 1981; OHNO and TANIGUCHI 1981; TAVERNIER et al. 1981; LAWN et al. 1981 c; GR et al. 1981). These chromosomal genes and their flanking 5′ and 3′ regions ha been extensively characterized.

All of the cloned chromosomal IFN-α and -β_1 genes lack introns. The chron somal DNA sequence is completely colinear with the nucleotide sequence of IF and -β_1 mRNA species. The absence of intervening sequences in these IFN-α a -β_1 genes is unusual in that eukaryotic genes (except the histones) contain "redu dant" noncoding DNA (introns) interspersed within the coding regions (exo (HAMER and LEDER 1979). The DNA sequence in the introns frequently diver much more rapidly than in the exons of related genes (HEILIG et al. 1980). Seve of the cross-hybridizing and closely related IFN-α genes are also closely linked tandem and contain inverted repeats in the flanking regions, suggestive of a ge duplication mechanism in the evolution of these IFN-α genes (NAGATA et al. 19 LAWN et al. 1981 a, b).

Figure 2 presents a comparison of the nucleotide sequences in the 5′ noncodi and the 5′ flanking region in several IFN-α genes and in the IFN-β_1 gene. The flanking region immediately upstream from an mRNA sequence is thought to pl an important role in the regulation of gene expression. Specific sequer homologies are present in the 5′ flanking region of eukaryotic structural gen These regions could represent loci where RNA polymerase II binds to the DN to initiate transcription or where other inducers of regulatory molecules could bi

to activate or repress the transcriptional acitivity of genes. Figure 2 reveals regions of distinct homology between the IFN genes, several of which are also seen in other eukaryotic genes and may thus be important in the regulation and expression of these genes.

a. There is a transcription initiation/capping site CA* approximately 70 nucleotides upstream from the translation initiation codon ATG.

b. A sequence TATTTAA approximately 31 nucleotides upstream from the presumed cap site is common to all the IFN-α genes. A similar sequence TATAAA 30 nucleotides upstream from the cap site is seen in the IFN-β_1 gene. This sequence is thought to play an important role in positioning the initiation of transcription and is generally found at a similar distance from the transcriptional start sites of eukaryotic genes (GOLDBERG 1979; GROSVELD et al. 1981; BAKER et al. 1981).

c. The sequence GAAAGT$^{A}_{G}$ is present at position -77 in the IFN-α genes and at -72 in the IFN-β_1 genes. This region presumably serves as a controlling or recognition region for transcription by RNA polymerase II (BENOIST et al. 1980; WASYLYK et al. 1980).

d. The sequence CTCTGAA (-57 to -51 in IFN-β_1) in present at about the same distance in chicken ovalbumin (BENOIST et al. 1980) and conalbumin (COCHET et al. 1979), and in modified form in the IFN-α genes.

e. Further upstream, the sequence TACTAAAATG is observed in IFN-β_1 and to some degree in the IFN-α genes. A similar sequence also occurs at a considerable distance (-125 to -140 nucleotides) from the cap site in human insulin, chicken ovalbumin, and chicken conalbumin genes (BENOIST et al. 1980; COCHET et al. 1979; BELL et al. 1980). Homologies indicated in e and c may be common to inducible proteins.

f. Small direct repeats are present several hundred (300) nucleotides upstream from some of the IFN-α genes and the IFN-β_1 gene in addition to a palindromic sequence in positions -280 to -240 (LAWN et al. 1981 a, b; GROSS et al. 1981).

These homologies in 5′ flanking sequence suggest not only that these genes may have evolved from a common ancestor, but that these regions may also have an important function in the expression of IFN genes. Multiple polyadenylation signals (AATAAA or ATTAAA) are seen in the 3′ flanking region of several IFN-α genes. It is clear that the same gene can give rise to mRNA species which utilize different poly(A) sites (GOEDDEL et al. 1981; LAWN et al. 1981 b).

F. Other Human IFN-α and -β Genes

Recent evidence suggests the existence of a second set of human IFN-α mRNAs which code for IFNs which are serologically of the α type but which do not cross-hybridize IFN-α_1-related cDNA probes, even under relaxed hybridization conditions (SAGAR et al. 1981; SEHGAL et al. 1981 a, b). This set of unusual mRNA species of length 1.6–3 kilobases (designated IFN-α_L) can be resolved from the conventional IFN-α mRNAs of length 0.7–1.4 kilobases (designated IFN-α_S) by electrophoresis of RNA through agarose–CH_3HgOH gels.

Similarly, electrophoresis of RNA through agarose–CH_3HgOH gels has led to the recognition of five distinct IFN-β mRNAs designated IFN-β_1 through IFN-β_5 (of lengths 0.9, 1.3, 1.8, 0.7, 0.9 kilobases, respectively) (SEHGAL and SAGAR 1980; SEHGAL et al. 1981 a; SAGAR et al. 1981, 1982). Even though these mRNAs code for IFNs which are serologically of the β type, their nucleic acids do not appear to cross-hybridize (SEHGAL and SAGAR 1980; WEISSENBACH et al. 1980; SAGAR et al. 1982). The molecular cloning of IFN-β_2 cDNA has been recently reported (WEISSENBACH et al. 1980). Thus, the IFN-α and -β gene family is even more complex than has been described in Sects. D and E. There appears to be even greater variability in IFN structural genes than had been anticipated.

G. Chromosomal Localization

Several of the IFN-α_S genes are closely linked and are arranged in tandem in chromosomal DNA (NAGATA et al. 1981; LAWN et al. 1981 a). Although most of these genes have been localized to human chromosome 9 (OWERBACH et al. 1981), it is unclear whether all of these are present on chromosome 9. The IFN-β_1 gene has also been localized to human chromosome 9 (MEAGER et al. 1979; OWERBACH et al. 1981). Nevertheless, the IFN-β_1 gene is not closely linked to the IFN-α_S genes since large (35–40 kilobases) segments of chromosomal DNA containing IFN-β_1 have been found to lack IFN-α sequences (GROSS et al. 1981).

The chromosomal localization of IFN-α_L genes (SEHGAL et al. 1981 a, b) is not known. It has been clearly shown that the other human IFN-β genes are widely dispersed in the human genome (TAN et al. 1974; SLATE and RUDDLE 1979, 1980; SAGAR et al. 1982). The available data are consistent with the localization of IFN-β_2 to human chromosome 5, and IFN-β_3 and -β_5 to chromosome 2 (SAGAR et al. 1982). In addition, there may exist another IFN-β on a chromosome other than 2, 5, or 9 (SAGAR et al. 1982). Although IFN-β_1 is a gene without introns, there is suggestive evidence that IFN-β_2 may be a gene with introns (SEHGAL and TAMM 1980; M. REVEL 1980, personal communication).

Several of the IFN-α_S genes localized to chromosome 9 are expressed in a coordinate manner (GOEDDEL et al. 1980 a, 1981). The IFN-β_1 gene which is also localized to chromosome 9 can be expressed independently of these α genes following poly(I)·poly(C) induction of diploid human fibroblasts (TANIGUCHI et al. 1979; 1980 a, b) as well as coordinately with the IFN-α genes following virus induction of human myeloblastoid cells (GOEDDEL et al. 1980 b). Furthermore, the various IFN-β genes can be expressed in a grossly noncoordinate manner in poly(I)·poly(C)-induced diploid human fibroblasts (SEHGAL and SAGAR 1980; SAGAR et al. 1982).

The expression of IFN-α_S genes can be inhibited and that of IFN-α_L genes enhanced when human peripheral blood leukocytes are induced with Sendai virus in the presence of 5,6-dichloro-1-β-D-ribofuranosylbenzimidazole (SEHGAL et al. 1981 b). It is likely that this variability in the expression of human IFN genes reflects the complex structural relationships between them. Insights into these phenomena may have to await the molecular cloning and characterization of the recently discovered human α and β genes. Similarly, human IFN-γ genes await detailed characterization.

H. IFN Structural Genes in Other Species

IFNs are expressed in a wide range of animal species (STEWART 1979). The murine IFN genes are likely to be as complex as the human IFN genes. Murine IFN-α, -β, and -γ have been clearly recognized (YAMAMOTO and KAWADE 1980; OSBORNE et al. 1979). The NH_2 terminal amino acid sequence of a species of murine IFN-α reveals good homology with a species of human IFN-α and that of murine IFN-β reveals some homology with human IFN-β_1 (Fig. 1; TAIRA et al. 1980). Antisera to human IFN-α cross-react with a species of murine IFN-α (STEWART and HAVELL 1980; HAVELL and CARTER 1981). DNA probes derived from human IFN-α genes appear to cross-hybridize with analogous sequences in the murine genome (OWERBACH et al. 1981). While the known human IFN-α proteins lack carbohydrate moieties and human IFN-β_1 contains carbohydrate, both murine IFN-α and -β proteins are glycoproteins (HAVELL and CARTER 1981). Appropriate differences (e.g., Asn-X-Ser or Asn-X-Thr sequences) between the structure of human IFN-α and murine IFN-α genes can be anticipated. It is likely that there will be rapid progress in the elucidation of the structure of not only the murine IFN genes, but of those in a wide variety of animal species.

J. Conclusions

IFNs represent a highly complex multigene family which consists of numerous closely related as well as several distantly related genes. Some of the genes which are closely related are present in a cluster on chromosome 9 (IFN-α_s genes) whereas several of the genes which are more distantly related are dispersed in the human genome (IFN-β genes). Although recent advances in the characterization of some human IFN genes have provided fascinating insights into some of the structural relationship that exist in this gene family, several of the newly recognized IFN genes still remain to be characterized. It is likely that even more exciting insights lie ahead.

Acknowledgments. We thank Dr. IGOR TAMM for numerous helpful discussions. Research in the authors' laboratory is supported by Grant AI-16262 from the NIAID. P. B. S. is the recipient of a Junior Faculty Research Award from the American Cancer Society and A. D. S. is supported by an NIH Institutionl predoctoral fellowship.

References

Allen G, Fantes KH (1980) A family of structural genes for human lymphoblastoid (leucocyte-type) interferon. Nature 287:408–411

Baker CC, Herisse J, Courtois G, Galibert F, Ziff E (1979) Messenger RNA for the Ad 2 DNA binding protein: DNA sequences encoding the first leader and heterogeneity at the mRNA 5′ end. Cell 18:569–580

Bell GI, Pictet RL, Rutter WJ, Cordell B, Tischer E, Goodman HM (1980) Sequence of the human insulin gene. Nature 284:26–32

Benoist C, O'Hare K, Breathnach R, Chambon P (1980) The ovalbumin genese-quence of putative control regions. Nucleic Acids Res 8:127–142

Cochet M, Gannon F, Hen R, Maroteaux L, Perrin F, Chambon P (1979) Organization and sequence studies of the 17 piece chicken conalbumin gene. Nature 282:567–574

Dayoff MO (1978) Atlas of protein sequence and structure, vol 5, suppl 3. National Biomedical Research Foundation, Washington, DC

Degrave W, Derynck R, Tavernier J, Haegeman G, Fiers W (1981) Nucleotide sequence of the chromosomal gene for human fibroblast (β_1) interferon and of the flanking regions. Gene 14:137–143

Derynck R, Content J, DeClercq E, Volckaert G, Tavernier J, Devos R, Fiers W (1980) Isolation and structure of a human fibroblast interferon gene. Nature 285:542–546

Derynck R, Remaut E, Saman E, Stanssens P, DeClercq E, Content J, Fiers W (1980b) Expression of human fibroblast interferon gene in Escherichia coli. Nature 287:193–197

Goeddel DV, Yelverton E, Ullrich A, Heyneker HL, Miozzari G, Holmes W, Seeburg PH, Dull T, May L, Stebbing N, Crea R, Maeda S, McCandliss R, Sloma A, Tabor JM, Gross M, Familletti PC, Pestka S (1980a) Human leucocyte interferon produced by E. coli is biologically active. Nature 287:411–416

Goeddel DV, Shepard HM, Yelverton E, Leung D, Crea R, Sloma A, Pestka S (1980b) Synthesis of human fibroblast interferon by E. coli. Nucleic Acids Res 8:4057–4074

Goeddel DV, Leung DW, Dull TJ, Gross M, Lawn RM, McCandliss R, Seeburg PH, Ullrich A, Yelverton E, Gray P (1981) The structure of eight distinct cloned human leucocyte interferon cDNAs. Nature 290:20–26

Goldberg M (1979) Ph. D. Thesis, Stanford University

Gross G, Mayr U, Bruns W, Grosveld F, Dahl HHM, Collins J (1981) The structure of a thirty-six kilobase region of the human chromosome including the fibroblast interferon gene IFN β. Nucleic Acids Res 9:2495–2507

Grosveld CC, Shewmaker CK, Jat P, Flavell RA (1981) Localization of DNA sequences necessary for transcription of the rabbit β globin gene in vitro. Cell 25:215–226

Hamer DH, Leder P (1979) Splicing and the formation of stable RNA. Cell 18:1299–1302

Havell EA, Carter WA (1981) Effects of tunicamycin on the physical properties and antiviral activities of murine L cell interferon. Virology 108:80–86

Heilig R, Perrin F, Cannon F, Mandel JL, Chambon P (1980) The ovalbumin gene family: structure of the X gene and evolution of duplicated split genes. Cell 20:625–637

Houghton M, Jackson IJ, Porter AG, Doel SM, Catlin GH, Barber C, Carey NH (1980) The absence of introns within a human fibroblast interferon gene. Nucleic Acids Res 9:247–266

Knight E Jr (1976) Interferon: Purification and initial characterization from human diploid cells. Proc Natl Acad Sci USA 73:520–523

Knight E Jr, Hunkapiller MW, Korant BD, Hardy RWF, Hood LE (1980) Human fibroblast interferon: amino acid analysis and amino terminal amino acid sequence. Science 207:525–526

Lawn RM, Adelman J, Dull TJ, Gross M, Goeddel D, Ullrich A (1981a) DNA sequence of two closely linked human leucocyte interferon genes. Science 212:1159–1162

Lawn RM, Gross M, Houck CM, Franke AE, Gray PV, Goeddel DV (1981b) DNA sequence of a major human leucocyte interferon gene. Proc Natl Acad Sci USA 78:5435–5439

Lawn RM, Adelman J, Franke AE, Houck CM, Gross M, Najarian R, Goeddel DV (1981c) Human fibroblast interferon gene lacks introns. Nucleic Acids Res 9:1045–1052

Levy WP, Rubinstein M, Shively J, Delvalle U, Lai CY, Moschera J, Brink L, Gerber I, Stein S, Pestka S (1981) Amino acid sequence of a human leucocyte interferon. Proc Natl Acad Sci USA 78:6186–6190

Maeda S, McCandliss R, Gross M, Sloma A, Familletti PC, Tabor JM, Evinger M, Levy WP, Pestka S (1980) Construction and identification of bacterial plasmids containing nucleotide sequence for human leukocyte interferon. Proc Natl Acad Sci USA 77:7010–7013

Mantei N, Schwarzstein M, Streuli M, Panem S, Nagata S, Weissman C (1980) The nucleotide sequence of a cloned human leucocyte interferon cDNA. Gene 10:1–10

Meager A, Graves M, Burke DC, Swallow DM (1979) Involvement of a gene on chromosome 9 in human fibroblast interferon production. Nature 280:493–495

Nagata S, Taira H, Hall A, Johnsrud L, Streuli M, Escodi J, Boll W, Cantell K, Weissmann C (1980a) Synthesis in E. coli of a polypeptide with human leucocyte interferon activity. Nature 284:316–320

Nagata S, Mantei N, Weissmann C (1980b) The structure of one of the eight or more distinct chromosomal genes for human interferon-α. Nature 287:401–408

Nagata S, Brack C, Henco K, Schamböck A, Weissmann C (1981) Partial mapping of ten genes of the human interferon-α family. J Interferon Res 1:333–336

Neuberger A, Gottschalk A, Marshall RD, Spiro RG (1972) In: Gottschalk A (ed) The glycoproteins: their composition, structure and function. Elsevier, Amsterdam, p 450

Ohno S, Taniguchi T (1981) Structure of a chromosomal gene for human interferon β. Proc Natl Acad Sci USA 78:5305–5309

Osborne LC, Georgiades JA, Johnson HM (1979) Large scale production and partial purification of mouse immune interferon. Infect Immun 23:80–86

Owerbach D, Rutter WJ, Shows TB, Gray P, Goeddel DV, Lawn RM (1981) Leukocyte and fibroblast genes are located on human chromosome 9. Proc Natl Acad Sci USA 78:3123–3127

Proudfoot NJ, Brownlee GG (1976) 3′ Noncoding region sequences in eukaryotic messenger RNA. Nature 263:211–214

Rubinstein M, Levy WP, Moschera JA, Lal GY, Hershberg RD, Bartlett RT, Pestka S (1981) Human leucocyte interferon: isolation and characterization of several molecular forms. Arch Biochem Biophys 210:307–318

Sagar AD, Pickering LA, Sussman-Berger P, Stewart WE II, Sehgal PB (1981) Heterogeneity of interferon mRNA species from Sendai virus-induced human lumphoblastoid (Namalva) cells and Newcastle disease virus-induced murine fibroblast (L) cells. Nucleic Acids Res 9:149–159

Sagar AD, Sehgal PB, Slate DL, Ruddle FH (1982) Multiple human β interferon genes. J Exp Med 156:744–755

Sehgal PB, Sagar AD (1980) Heterogeneity of poly(I)·poly(C)-induced human fibroblast interferon mRNA species. Nature 287:95–97

Sehgal PB, Tamm I (1980) The transcription unit for poly(I)·poly(C)-induced human fibroblast interferon messenger RNA. Virology 102:245–249

Sehgal PB, Sagar AD, Braude IA, Smith D (1981a) Heterogeneity of human α and β interferon mRNA species. In: DeMaeyer E, Schellekens H (eds) The biology of the interferon system, North-Holland/Elsevier, Amsterdam, p 43

Sehgal PB, Sagar AD, Braude IA (1981b) Further heterogeneity of human α interferon mRNA species. Science 214:803–805

Sehgal PB, Pfeffer LM, Tamm I (1982) Interferon and its inducers. In: Came PE, Caliguiri LA (eds) Chemotherapy of viral infections. Hdbk Exp Pharm 61:205–311 Springer Berlin, Heidelberg, New York

Shepard M, Leung D, Stebbing N, Goeddel DV (1981) Synthesis in E. coli of a naturally occurring mutant fibroblast interferon. Abstr of the 5th international congress of virology, Strasbourg, p 89

Slate DL, Ruddle FH (1979) Fibroblast interferon in man is coded by two loci on separate chromosomes. Cell 16:171–180

Slate DL, Ruddle FH (1980) Somatic cell genetic analysis of interferon production and response. Ann NY Acad Sci 350:174–178

Soreq H, Sagar AD, Sehgal PB (1981) Translational activity and functional stability of human fibroblast β_1 and β_2 interferon mRNAs lacking 3′-terminal RNA sequences. Proc Natl Acad Sci USA 78:1741–1745

Stewart WE II (ed) (1979) The interferon system. Springer Berlin Heidelberg New York

Stewart WE II, Havell EA (1980) Characterization of a subspecies of mouse interferon cross-reactive on human cells and antigenically related to human leukocyte interferon. Virology 101:315–318

Streuli M, Nagata S, Weissmann C (1980) At least three human type α interferons: Structure of α2. Science 209:1343–1347

Streuli M, Hall A, Boll W, Stewart WE II, Nagata S, Weissmann C (1981) Target cell specificity of two species of human interferon-α produced in Escherichia coli and of hybri molecules derived from them. Proc Natl Acad Sci USA 78:2848–2852

Taira H, Broeze RJ, Jayaram BM, Lengyel P, Hunkapillar MW, Hood LE (1980) Mou interferons: Amino terminal amino acid sequences of various species. Science 207:52 529

Tan YH (1974) The somatic cell genetics of human interferon: Assignment of human inte feron loci to chromosomes 2 and 5. Proc Natl Acad Sci USA 71:2251–2255

Tan YH, Barakat F, Berthold W, Smith-Johannsen H, Tan C (1979) The isolation ar amino acid/sugar composition of human fibroblastoid interferon. J Biol Che 254:8067–8073

Taniguchi T, Sakai M, Fujii-Kuriyama Y, Muramatsu M, Kobayashi S, Sudo T (1979) Co struction and identification of a bacterial plasmid containing the human fibroblast i terferon gene sequence. Proc Jpn Acad Ser B 55:464–469

Taniguchi T, Ohno S, Fujii-Kuriyama Y, Muramatsu M (1980a) The nucleotide sequen of human fibroblast interferon cDNA. Gene 10:11–15

Taniguchi T, Guarente L, Roberts TM, Kimelman D, Douhan J, Ptashne M (1980b) E pression of the human fibroblast interferon gene in Escherichia coli. Proc Natl Acad S USA 77:5230–5233

Taniguchi T, Mantei N, Schwarstein M, Shigekazu N, Muramatsu M, Weissmann C (198 c) Human leucocyte and fibroblast interferons are structurally related. Nature 285:54 549

Tavernier J, Derynck R, Fiers W (1981) Evidence for a unique human fibroblast interfero (IFN β_1) chromosomal gene, devoid of intervening sequences. Nucleic Acids Res 9:461 471

Wasylyk B, Derbyshire R, Guy A, Molko D, Roget A, Tedule R, Chambon P (1980) Spe cific in vitro transcription of conalbumin gene is drastically decreased by single-poir mutations in T-A-T-A box homology sequence. Proc Natl Acad Sci USA 77:7024–7028

Weissenbach J, Chernajovsky Y, Zeevi M, Shulman L, Soreq H, Nir U, Wallach D, Perr caudet M, Tiollais P, Revel M (1980) Two interferon mRNAs in human fibroblasts: I vitro translation and Escherichia coli cloning studies. Proc Natl Acad Sci USA 77:7152 7156

Wetzel R (1981) Assignment of the disulphide bonds of leucocyte interferon. Natu 289:606–607

Yamamoto Y, Kawade Y (1980) Antigenicity of mouse interferons: Distinct antigenicity the two L cell interferon species. Virology 103:80–88